

1979

# A Study of the Relationship of Pupil Achievement to the Degree of Teacher Implementation of an Individualized Elementary Reading Program.

Patience Weidt Keisler

*Louisiana State University and Agricultural & Mechanical College*

Follow this and additional works at: [https://digitalcommons.lsu.edu/gradschool\\_disstheses](https://digitalcommons.lsu.edu/gradschool_disstheses)

---

## Recommended Citation

Keisler, Patience Weidt, "A Study of the Relationship of Pupil Achievement to the Degree of Teacher Implementation of an Individualized Elementary Reading Program." (1979). *LSU Historical Dissertations and Theses*. 3444.  
[https://digitalcommons.lsu.edu/gradschool\\_disstheses/3444](https://digitalcommons.lsu.edu/gradschool_disstheses/3444)

This Dissertation is brought to you for free and open access by the Graduate School at LSU Digital Commons. It has been accepted for inclusion in LSU Historical Dissertations and Theses by an authorized administrator of LSU Digital Commons. For more information, please contact [gradetd@lsu.edu](mailto:gradetd@lsu.edu).

## **INFORMATION TO USERS**

**This was produced from a copy of a document sent to us for microfilming. While the most advanced technological means to photograph and reproduce this document have been used, the quality is heavily dependent upon the quality of the material submitted.**

**The following explanation of techniques is provided to help you understand markings or notations which may appear on this reproduction.**

- 1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting through an image and duplicating adjacent pages to assure you of complete continuity.**
- 2. When an image on the film is obliterated with a round black mark it is an indication that the film inspector noticed either blurred copy because of movement during exposure, or duplicate copy. Unless we meant to delete copyrighted materials that should not have been filmed, you will find a good image of the page in the adjacent frame.**
- 3. When a map, drawing or chart, etc., is part of the material being photographed the photographer has followed a definite method in "sectioning" the material. It is customary to begin filming at the upper left hand corner of a large sheet and to continue from left to right in equal sections with small overlaps. If necessary, sectioning is continued again—beginning below the first row and continuing on until complete.**
- 4. For any illustrations that cannot be reproduced satisfactorily by xerography, photographic prints can be purchased at additional cost and tipped into your xerographic copy. Requests can be made to our Dissertations Customer Services Department.**
- 5. Some pages in any document may have indistinct print. In all cases we have filmed the best available copy.**

**University  
Microfilms  
International**

**300 N. ZEEB ROAD, ANN ARBOR, MI 48106  
18 BEDFORD ROW, LONDON WC1R 4EJ, ENGLAND**

KEISLER, PATIENCE WEIDT

A STUDY OF THE RELATIONSHIP OF PUPIL ACHIEVEMENT TO THE  
DEGREE OF TEACHER IMPLEMENTATION OF AN INDIVIDUALIZED  
ELEMENTARY READING PROGRAM

The Louisiana State University and Agricultural and Mechanical Col.

PH.D. 1979

University  
Microfilms  
International

300 N. Zeeb Road, Ann Arbor, MI 48106

18 Bedford Row, London WC1R 4EJ, England

A STUDY OF THE RELATIONSHIP OF PUPIL ACHIEVEMENT  
TO THE DEGREE OF TEACHER IMPLEMENTATION OF AN  
INDIVIDUALIZED ELEMENTARY READING PROGRAM

A Dissertation

Submitted to the Graduate Faculty of the  
Louisiana State University and  
Agricultural and Mechanical College  
in partial fulfillment of the  
requirements for the degree of

Doctor of Philosophy

in

The Interdepartmental Program of Education

by

Patience W. Keisler  
B.S., Midland College, 1952  
M.Ed., Louisiana State University, 1972

December, 1979

## ACKNOWLEDGEMENTS

The writer hereby acknowledges with gratitude the indispensable help and unique contributions of the many people, named and unnamed, who made possible the completion of her graduate program. For their major roles in her graduate study, the writer expressly thanks:

Dr. William M. Smith, the writer's first graduate school professor, who gave her the courage to launch into more than one graduate program.

Dr. Helen M. Cookston, her major professor, who has been both patient and impatient at the right times while giving generously of herself to advance this study.

Dr. Doris J. Conway, whose probing, analytical thinking, intense concern, and high aspiration for her students have challenged the writer.

Dr. Sam Adams, who could always point out just the right formula and function to fit the problem.

Dr. Eric L. Thurston, whose jovial harrassment communicated concern and confidence.

Dr. Jay D. Edwards, who guided the writer into the fascinating worlds of cultural and linguistic studies and introduced her to the structuralist view of literature.

Dr. Barbara M. Strawitz, Dr. Prentiss E. Schilling, and Dr. James W. Firnberg, who graciously gave time and

attention to help the writer learn and apply the processes of designing, conducting, and analyzing research.

Grateful acknowledgement is accorded two valued colleagues, Mary Ellen Jordan and Dr. Molly Newkome, for their highly respected counsel and, in particular, for their careful attention to collection of data.

The writer also acknowledges the vision and leadership of Dr. Donald L. Hoover and Mrs. Edna West, and the enthusiasm of the capable ERIP teachers who worked diligently to use and share ever more effective ways of enticing children to read.

To Julia Craige, who freed the writer for many hours which would otherwise not have been available for educational and professional endeavors, and to many loving friends who have been generously kind and helpful even when they may not have understood, the writer is indeed indebted.

Most of all, the writer expresses profound appreciation to her husband and all-time favorite professor, James, for lending his objective viewpoint and sensitive understanding, and for freely giving his constant loving support, and to her sons, Jim, Billy-Bryan, and Paul, for their good-humored spirit of tolerance, positive intellectual support, and active help at home and school in innumerable ways. To these four cherished people, the writer acknowledges with pleasure and pride that their high standards of academic endeavor and logical reasoning are a constant inspiration.

## TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS . . . . .	ii
LIST OF TABLES . . . . .	vii
LIST OF FIGURES. . . . .	ix
ABSTRACT . . . . .	x
Chapter	
1. INTRODUCTION. . . . .	1
Statement of the Problem. . . . .	2
Questions to be Answered. . . . .	2
Importance of the Study . . . . .	3
Delimitation of the Study . . . . .	5
Definition of Terms . . . . .	6
Design of the Study . . . . .	7
Organization of the Study . . . . .	9
2. REVIEW OF SELECTED LITERATURE . . . . .	10
Teacher Effectiveness Research. . . . .	10
Systematic Classroom Observation. . . . .	11
Design of Teaching Models . . . . .	11
Presage Variables . . . . .	11
Context Variables . . . . .	12
Process Variables . . . . .	13
Product Variables . . . . .	14
Development and Use of Observation Measures. . . . .	14
Summary . . . . .	22

	Page
Relevant Studies . . . . .	22
Importance of Process-Product Studies. .	22
Flanders' Word . . . . .	23
Documentation of Implementation. . . . .	25
Variety of Approach. . . . .	28
Teaching of Skills . . . . .	45
Classroom Interaction. . . . .	55
Summary. . . . .	63
Summary. . . . .	64
3. DESIGN AND PROCEDURES. . . . .	65
Design of the Study. . . . .	65
Development of the Observation Checklist .	66
Rationale for the Use of a Sign System .	66
Content of the Checklist . . . . .	67
Validation Procedures. . . . .	69
Reliability Studies. . . . .	71
Modification of the Recording Form . . .	72
Selection of the Stratified Sample . . . .	75
Collection of Observation Data . . . . .	80
Administration of Achievement Tests. . . .	82
Treatment of the Data. . . . .	83
Summary. . . . .	83
4. PRESENTATION AND ANALYSIS OF DATA. . . . .	85
Reading Measures and Implementation	
Measures . . . . .	86



	Page
Effects of Variables on Reading Achievement . . . . .	89
Summary . . . . .	99
5. SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS .	100
Summary . . . . .	100
Findings. . . . .	102
Conclusions . . . . .	104
Recommendations . . . . .	106
SELECTED BIBLIOGRAPHY. . . . .	109
APPENDICES . . . . .	118
A. Design of the Elementary Reading Improvement Program . . . . .	118
B. Observation Checklist and Classroom Tally Sheet. . . . .	122
C. Communications with Elementary Reading Improvement Program Personnel . . . . .	127
D. Checklist Form Used in Validation Study . . . . .	133
E. School Board Permission . . . . .	137
F. Instructions for Administration of Achievement Tests . . . . .	141
G. Printout of Primary Data. . . . .	145
H. Schools of the Study. . . . .	147
I. Gains versus Unadjusted Pretest Scores by Implementation Type. . . . .	149
VITA . . . . .	151

## LIST OF TABLES

Table	Page
1. Reliability Studies: Total Observation Checklist Scores for Three Observers Independently Scoring the Same Classroom Simultaneously . . . . .	73
2. Checklist Scores for Initial Observation of Highest and Lowest Implementing Classrooms in Each Five-School Cluster . . .	76
3. Distribution of Classes. . . . .	78
4. Distribution of Classrooms in Sample by School, Type, and Grade. . . . .	79
5. Correlation Coefficients Showing Relationships Among Checklist Scores and Among Reading Achievement Measures for Total Sample . . . . .	87
6. Frequency Distribution of Classrooms According to Mean Pretest and Mean Gain for Total Sample. . . . .	88
7. SRA Reading Achievement in Raw Score Measures for High and Low Implementing Classrooms . . . . .	89
8. SRA Reading Achievement in Raw Score Measures Classified by Student Race. . . . .	92
9. SRA Reading Achievement in Raw Score Measures Classified by Student Sex . . . . .	92
10. SRA Reading Achievement in Raw Score Measures Classified by Student Sex by Student Race. . . . .	93
11. SRA Reading Achievement in Raw Score Measures Classified by Grade Level and Implementation Type. . . . .	93
12. SRA Reading Achievement in Raw Score Measures Classified by Teacher Race and Implementation Type. . . . .	96

Table		Page
13.	SRA Reading Achievement in Raw Score Measures Classified by Student and Teacher Race . . . . .	97
14.	Frequency Distribution of Classrooms According to Mean Pretest and Mean Gain for High Implementers . . . . .	150
15.	Frequency Distribution of Classrooms According to Mean Pretest and Mean Gain for Low Implementers. . . . .	150

## LIST OF FIGURES

Figure	Page
1. Comparison of Simultaneous Observation Checklist Records of the Three Observers. . .	74
2. Selection of Sample Based on Checklist Scores for Initial Observation in 118 Classrooms Meeting the Criteria . . . . .	77
3. Comparison of Mean Gains of High and Low Implementers by Grade Level . . . . .	95

## ABSTRACT

This study is an evaluation of the Elementary Reading Improvement Program (ERIP), an individualized, multiapproach program in reading-language arts instruction in the elementary grades. An Observation Checklist was developed and validated for use in assessing the degree of classroom implementation of the program. The checklist directed observation to three aspects of implementation: variety of approaches to reading instruction, diagnosis and development of specific reading skills, and teacher-pupil interaction. Gains in pupil reading achievement were used to determine the effects of the variables: degree of implementation, pupil sex, pupil race, teacher race, and grade level.

The research was designed to compare classroom process (the degree of implementation of the program) with learning product (gains shown in reading achievement test scores). Study of observation instruments preceded the development and validation of the sign type Observation Checklist for use in this study. The three instructional consultants in the ERIP participated in observer training and reliability studies. Stratified samples of the highest and lowest implementing classes in each five-school cluster were selected to serve as the treatment and control groups, respectively. The sample consisted of forty-three teachers and 994 students.

To provide the process data, a total of three observations were completed in each classroom of the sample group. The mean of the total checklist scores for the three observations was the index of implementation for a classroom.

Product data consisted of reading sections of the SRA Achievement Test Series. These standardized achievement tests were administered by the classroom teachers who had been provided with written instructions prior to the pre- and post-testing periods.

Process data (the checklist scores) and product data (the reading score gains) both used the class as the unit of measure in evaluating this reading program. Analyses of variance were used to establish the effects of the variables on gains in reading achievement using test scores adjusted for initial differences. Correlation coefficients determined levels of significance of the variances between results for the groups under comparison.

The findings of this study indicated that the Observation Checklist scores of Parts I, II, and III, and the total correlated to a highly significant degree ( $p < .01$ ). As a result, findings could be discussed in terms of the index of implementation. Pretest and posttest mean scores had a highly significant correlation ( $p < .01$ ). Therefore, pretest scores could be considered good predictors of posttest levels. The amount of gain in reading achievement was negatively correlated, to a highly significant degree ( $p < .01$ ), with mean

pretest scores. Low implementing classes of the ERIP showed significantly greater ( $p < .05$ ) raw mean gains in reading achievement than did the high implementers. Thus, according to this data, the ERIP was not advantageous. Pupil race, grade level, and the interaction between grade level and implementation type appeared to make highly significant differences ( $p < .01$ ) in pupil reading achievement gains. Study of the data suggested no discernable explanation for the sharp contrasts among implementation types and grades. The data demonstrated no significant differences in reading gains by the variables of pupil sex nor teacher race, nor by the interactions between pupil race and pupil sex, between teacher race and implementation type, nor between teacher race and pupil race.

## Chapter 1

### INTRODUCTION

Today's democratic society requires the best possible development of each individual's language skills. Lavatelli (1973) explaining Piaget, Bruner (1964), and Dewey (1961), among others, stressed the idea that learning is change that occurs in the individual as a result of some new understanding. Since learning occurs within the individual, it must be accomplished by the learner. The teacher can merely help by providing a setting designed to stimulate interest, suggesting alternatives, providing--in appropriate contexts--instruction and practice in specific skills, and supporting and encouraging the learner to develop independence in learning.

To meet the diverse needs of learners, teachers must be able to use many strategies. The Elementary Reading Improvement Program (ERIP) of the East Baton Rouge Parish School Board, Baton Rouge, Louisiana, was designed in response to desires expressed by elementary teachers for help in improving the effectiveness of reading instruction in their classrooms (Appendix A). Teachers were having difficulty meeting the wide range of individual pupil needs in self-contained classrooms. The main function of the ERIP was its in-service education to help teachers become more eclectic



and more effective in responding to pupils' differing needs and interests. If a teacher employed a wide variety of methods, materials, and learning inducements, according to pupil needs in the classroom (thus implementing the ERIP to a high degree), the pupils should be better enabled to make academic progress and develop the skills of logical thinking and decision making so necessary to citizens in a democratic society.

### Statement of the Problem

The problem of this study was to determine how pupil reading achievement was related to the degree of teacher implementation of the Elementary Reading Improvement Program of the East Baton Rouge Parish School Board, Baton Rouge, Louisiana.

### Questions to be Answered

The specific questions investigated were:

1. How was pupil reading achievement related to the classroom practices utilized in variety of approaches to reading instruction?
2. How was pupil reading achievement related to the classroom practices utilized in diagnosis and development of specific reading skills?
3. How was pupil reading achievement related to classroom interaction?

4. How was pupil reading achievement, with respect to pupil sex, pupil race, teacher race, and grade level, related to the index of implementation?

#### Importance of the Study

This study will provide a means of evaluation of the ERIP of the East Baton Rouge Parish School Board to determine whether this local program improved the teaching and learning of reading. Permission was secured from the Board to conduct the study (Appendix E). Many program evaluations looked at the initial and final status of the pupils studying under a particular curriculum package without looking at the process of teaching (Stake, 1967). Medley and Mitzel (1963) remarked:

[Typically] . . . the research worker limits himself to the manipulation or study of antecedents and consequents of whatever happens in the classroom . . . but never once looks into the classroom to see how the teacher actually teaches (p. 247).

Even as recently as 1970, Rosenshine (1970b) wrote:

Compared to the large number of descriptive studies, there have been relatively few studies of the relationship between measures obtained by use of observational systems and measures of class achievement adjusted for initial aptitude or ability (p. 293).

This present study, however, included not only pre- and post-tests of pupils, but also used data collected in direct observation of classroom instruction. Since the study considered both the instructional process and the resultant educational product in terms of pupil achievement, it was a process-product study.

This study may have wider application than merely the ERIP. The schools in the ERIP were located in city as well as outlying areas thus serving both urban and rural families. The patrons of these schools represented a broad spectrum of socioeconomic levels. The racial makeup of the schools ranged from predominantly white to predominantly black. If this study indicated that the ERIP was successful in effecting reading gains in representative schools of Baton Rouge, the program may provide a useful model for other city school systems facing similar reading problems.

The ERIP was not limited to one package of materials but was based on a philosophy of individualizing instruction by employing different approaches for different learner needs, utilizing approved basal readers for teaching reading skills, and using a rich variety of children's literature to arouse interest and stimulate pupil motivation. The checklist could apply to other programs having a similar philosophy and employing a broad variety of reading materials.

The checklist items developed for use in this study could also be used for faculty or individual self-study in systematic teaching improvement programs or to guide student teaching-observers in what to look for in individualized learning situations, especially those pertaining to elementary reading and language arts.

Another notable facet of this study is that the instructional program under investigation employed various approaches in the same classroom. This design was in contrast

to many studies which compared classrooms stressing one approach with classrooms emphasizing a different approach or curriculum package.

#### Delimitation of the Study

The population for this study included classrooms of grades three through six in the fifteen public schools which were in the ERIP (Appendix H). A stratified sample of forty-eight classrooms was selected. Twenty-four of the highest implementers of the ERIP were considered the treatment group. Twenty-four of the lowest implementers became the control group. These forty-eight classrooms included 1,024 students who remained members of their respective classes from the pre-testing in September, 1975, through the posttesting in May, 1976.

The degree of teacher implementation, representing the instructional process data, was established on the basis of the Observation Checklist developed by the researcher for this purpose (Appendix B). The checklist was used to record observations by the three instructional consultants working in the classrooms to which each was regularly assigned in the ERIP.

The instruments used to measure reading achievement, representing the instructional product data, were Science Research Associates (SRA) Achievement Test Series in Reading, administered by the classroom teachers under the supervision of the guidance personnel of the school system.

Socioeconomic information on the student population was not available to the researcher.

### Definition of Terms

Cluster--A group of five schools in the ERIP composed of a center school and four nearby schools that worked together.

Center school--A school serving as a model teaching center, providing a location for the processing and circulation of media and print materials, and serving as headquarters for the instructional consultant in reading.

Instructional consultant--A specialist to help teachers individualize reading instruction, one consultant to each five-school cluster.

In-service education--A varied program of helping teachers individualize instruction. (The instructional consultant helped plan, arrange, and carry out faculty studies, work-study visits in other teachers' classrooms, pre- and post-visit conferences with individual teachers, lectures or workshops involving nationally known authorities, and provision of professional literature and classroom resources, both print and non-print.)

Work-study visit--A teacher, released from her class by a qualified substitute for a full day, observing and working in a classroom that used techniques and/or materials she was learning to use with her own pupils. (The instructional consultant planned with the visitor and visitee prior to the visit and did follow-up planning with the visitor for carry-over from the visited classroom to her own.)

Index of Implementation--The mean score computed from the total Observation Checklist scores made in one classroom and used as a measure of the degree of implementation of the ERIP.

Basal reader approach--Based on a coordinated, graded series of reading textbooks designed to give precise structure to sequential development of reading skills in a context of stories devised to capture children's interest.

Individualized approach--Involved each child working at his own pace with reading materials self-selected to fit his interests and reading abilities and reviewed in reading conferences with his teacher to evaluate progress and diagnose difficulties of the pupil.

Language experience approach--Used field trips, classroom activities, and personal experiences to provide the stimulus and content for language activities, beginning with the language skills the child had already acquired in listening, speaking, reading, and writing.

### Design of the Study

On the premise that teacher behavior affects pupil behavior, the ERIP aimed to increase teacher practices thought to improve pupil reading achievement. In order to determine the effectiveness of the ERIP, the teaching-learning process in the classroom was compared with the changes in reading achievement level of the pupils in that particular classroom.

The degree of implementation of the ERIP was established on the basis of the Observation Checklist developed for the purpose. The checklist measured classroom practices regarding variety of approach to teaching reading, diagnosis and direct teaching of specific reading skills, and classroom interaction. Instructional consultants, using the checklist, recorded observations of classroom activities in each of the third, fourth, fifth, and sixth grade classrooms which met the criteria for this study. On the basis of this initial observation, the classrooms were ranked according to degree of implementation. The twenty-four classrooms scoring highest

and the twenty-four classrooms scoring lowest became the treatment and control groups respectively. In these forty-eight classrooms, the instructional consultants recorded additional observations, making a total of three observations per classroom. The average of the three Observation Checklist scores was used as the index of implementation for each classroom.

Reading sections of the SRA Achievement Test Series were administered to all classes in grades three through six in the ERIP. In third and fourth grades, the SRA Primary II, Form E for pretest and Form F for posttest were used. In fifth and sixth grades, the SRA Multilevel, Form C for pretest and Form D for posttest were used. Pretests were administered in the week of September 8, 1975, and posttests, between May 4 and 14, 1976. Changes in reading achievement levels were determined by comparing pupils' posttest scores with pretest scores adjusted for initial ability on these standardized tests.

For each one of the forty-eight classrooms in the study, the mean change in pupil reading achievement level, with respect to pupil sex, pupil race, teacher race, and grade level, was compared with the index of implementation.

Analyses of variance procedures were applied by computer to the data to determine measures of the variables, and correlation coefficients were used to determine which variables exhibited significant relationships.

### Organization of the Study

The study was divided into five chapters: the background and introductory information were presented in Chapter 1; the review of related literature was the substance of Chapter 2; the development of the checklist and the procedures used in obtaining the data were traced in Chapter 3; a presentation and analysis of the data were included in Chapter 4; and conclusions and recommendations comprised Chapter 5.



## Chapter 2

### REVIEW OF SELECTED LITERATURE

Since the present study dealt with the analysis of the teaching-learning process as applied to reading and language arts, three areas of literature were consulted. First, the history of teacher effectiveness studies provided insights into the mercurial nature of the instructional process as a measurable entity. Next, a study of the systematic observation movement provided the design of a teaching model and gave guidance for the formulation of an observation checklist. Third, research literature involving correlational and implementational studies was consulted. Findings correlating pupil achievement (product) with classroom instruction (process), especially in elementary language arts, were investigated.

#### Teacher Effectiveness Research

The history of teacher effectiveness research indicated that there have been various attempts to use direct observation to find the relation between teacher behaviors and pupil change. The researchers consulted had been unable to isolate from the complex web of classroom activities a specific, critical factor that discriminated between effective and ineffective teachers.

### Systematic Classroom Observation

Despite the long history and considerable effort of numerous educators and researchers, the literature on teacher effectiveness seemed to produce little consistent or helpful information for predicting, identifying, or deliberately producing the effective ingredients of good teaching (Soar, 1970). However, about 1960, several changes were occurring and exerting influence on the research in teacher effectiveness. Important paradigms which described teaching models were designed and became widely accepted (Bennett, 1976). Observation schedules which employed objective measures and which directed attention to pupil and teacher interactions were developed (Rosenshine, 1973). Relationships among different aspects of the teaching model were examined.

#### Design of Teaching Models

A model for the study of classroom teaching was proposed by Mitzel in an unpublished report in 1957, and cited by Bennett (1976). Dunkin and Biddle's (1974) paradigm, an adaptation of Mitzel's model, showed four sets of variables involved in teaching: presage, context, process, and product variables.

Presage variables. Presage variables usually related to teacher background. Factors of presage included social class, age, education, intelligence, and personality of the teacher, as well as types of teaching experience and teaching

skills of the teacher. Many of the studies early in the century concentrated on finding, among these presage variables, the indicators of good teaching. Even much later, when Medley and Mitzel (1958) were studying the relationship between teacher personality and pupil growth, the researchers were still finding it very difficult to secure objective measures of teacher personality. Medley and Mitzel recommended that the observer should be provided with a form which listed particular teacher cues and an accompanying observation recording system. Flanders (1969) cited reviews by Howsam (1960) and Fattu (1962) which reported that the research had been unable to identify any teacher characteristic as a significant predictor of teacher effectiveness. Slight positive correlations between college grades and teacher effectiveness were probably due to their common basis of intelligence. Professional knowledge, such as that measured by the National Teacher Examination, has been somewhat more consistent as a predictor of good teaching performance. However, reviewers of the research conclude that teacher traits did not seem to provide dependable predictors of teaching effectiveness (Simon and Boyer, 1967).

Context variables. Context variables referred to teaching conditions. Some of the variables included in the context group were socioeconomic status (SES); ability and attitudes of pupils; school-community relations; school climate and size; and the class size, teaching equipment, and

instructional materials available in a specific classroom. A number of studies have searched this context group for significant factors. Simon and Boyer (1967) classified variables somewhat differently:

Settings and administrative arrangements (team teaching, small or large class structures) . . . are considered part of input (p. 17).

Of the use of these variables in research, Simon and Boyer (1967) reported:

Prior to the 1960's, almost all research on effective teaching concentrated on seeking links between characteristics of teachers or of teaching settings (input) and various kinds of pupil growth (output) (p. 16).

Process variables. The process variables were found in classroom teaching-learning activities, attitudes, and interactions. According to Simon and Boyer's definition, "Only interaction patterns between pupil and teacher are considered as the 'process' (1967:17)." Mitzel felt that study of the process variables of his model would prove to be of the most help in instructional research (Bennett, 1976). According to Simon and Boyer (1967):

Inclusion of process measures of teacher behavior in studies of teacher effectiveness has constituted a major change in this field. Data from these measures of what teachers and pupils 'do' in the classroom, as contrasted with what they 'have' or what they 'are' has contributed both to encouraging research results and a feeling of cautious optimism among writers in the field about the potential for building a viable theory of instruction with potential for implementation in practice (p. 16).

It was this group of process variables that were studied when classroom observation was employed.

Product variables. Product variables were the cognitive and/or affective pupil changes that resulted from the classroom activities. The outcomes measured, such as changes in pupil achievement level, were thought by some evaluators to be "the ultimate criteria for research on teacher effectiveness (Bennett, 1976:17)." However, Medley and Mitzel (1963) and Rosenshine and Furst (1973) indicated that teacher effectiveness could best be determined by studying the correlation of both observed process variables and product criteria. Since the present study related the observation of the classroom learning process to the pupil achievement product, it would be called a process-product study by some and a correlational study by others.

#### Development and Use of Observation Measures

In order to overcome inconsistencies of the early studies of teacher effectiveness reviewed by Morsh and Wilder (1954) and Ackerman (1954), the use of systematic observation techniques has been suggested. Accordingly, teaching behavior was to be documented in the most objective terms possible in order to provide data which could be related to learning outcomes. To develop dependable data regarding teacher effectiveness and the causes of pupil change, Ackerman (1954) stated the need for low inference observation and recording

of classroom activities. Numerous researchers shared Ackerman's concerns regarding the need for systematic classroom observation (Flanders, 1969). "Certainly there is no more obvious approach to research on teaching than direct observation of teachers while they teach (Medley, 1963:247)." Grannis (1972), commenting on the dearth of classroom behavior data, declared that collecting such information was requisite to research on how education affected children. Medley and Mitzel (1963) defined observational techniques as:

. . . procedures which use systematic observation of classroom behavior to obtain reliable and valid measurements of differences in the typical behaviors which occur in different classrooms, or in different situations in the same classroom. . . . The validity of measurements of behavior [depends on] . . . three conditions: 1. A representative sample of the behaviors to be measured must be observed. 2. An accurate record of the observed behaviors must be obtained. 3. The records must be scored so as to faithfully reflect differences in behavior (p. 250).

To study the teaching-learning process, it was necessary to specify what behaviors, activities, and other factors were hypothesized as being markers or indicators of good instruction. Since it was impossible to observe or record everything, only behaviors deemed relevant to the purpose of the study were abstracted from the total scene observed. These phenomena were stated in terms of the specific observable behaviors to be studied and were systematically recorded as observed.

One of the important aims of systematic observation was the devising of low inference measures. To be low

inference, items had to require, and permit, minimal observer judgment. The observer role was to record, not to evaluate (Medley, 1963). The less processing, analyzing, and/or estimating the observer did during the recording, the lower inference the measuring instrument was said to be and the more objective it was assumed to be. Low inference data remained as close as possible to the actual behavior. A somewhat general item could be reduced in inference by providing some specific examples which helped to define the items. Inference could also be lowered by protocols or conventions agreed upon by the designers and users of the observation instrument.

In addition to the careful selection and precise stating of items for a checklist, the organization of items on a recording form had to be considered. To be an effective research tool, the observation schedule or record had to be designed to facilitate reliable, accurate recording of observations. It also had to permit the accurate communication to others of what happened in the classroom situation observed so that a wider audience could benefit from what the researcher learned (Wang, 1973). Therefore, much effort of the systematic observation movement was applied to the complex task of analyzing the instructional process by constructing observation schedules which would provide data on specific aspects of pupil behavior, teacher behavior, or pupil-teacher interaction (McNeil, 1973; Furst, 1971; Medley, 1963).

To produce an accurate measure, an observation system had to possess the potential for interobserver reliability:

different observers independently recording highly similar responses to the same situation simultaneously. Therefore, the description, classification, or tally of instructional activities to be recorded had to be framed in accurate, unambiguous, objective terms.

In addition, the observers had to be trained to recognize and classify activities objectively and reliably according to a common interpretation of the criteria in the observational instrument. Observer training and guidelines for observation usually involved defining the parameters of the categories to be recorded. Deciding in which category to record an observed behavior was the main judgment to be employed by observers. The observers were to be carefully trained in the identification of each item or category and, when necessary, some arbitrary decisions were made and stated in a coders' manual about classifying certain behaviors. Thus, coders' manuals helped to achieve the purpose of direct observation: to secure an accurate, objective record of instructional activities (Rosenshine, 1971a).

According to Medley and Mitzel (1963), once a representative sample of behaviors had been selected and the means for accurate recording had been provided, a third condition also had to be met to insure validity: a system of scoring which could project an accurate picture of the situations observed and could make clear the differences between them. Three general types of scales were used: rating systems,



category systems, and sign systems (Rosenshine, 1973). Since the early 1960's, numerous instruments have been designed for use in the observation of classroom instruction. Many more rating systems have been developed for use in observing school programs, teachers, or learning environments, for evaluating student teachers, and for student rating of college courses and teachers than category systems. However, since Flanders' (1965) Interaction Analysis process became widely known, category systems have also increased in number. Observation systems were formerly classified as low inference category systems or high inference rating systems. However, ideas from both types have been integrated into some instruments. Thus, observation instruments can no longer be neatly pegged on those distinctions. Rosenshine and Furst (1973) suggested classifying observation instruments according to these three distinguishing characteristics: "the recording procedure, the scope and specificity of items, and the format used to code individual events (p. 132)."

In recording procedures, if an observed event or behavior was to be recorded each time it was seen, the observation instrument was called a category system. If an event was to be recorded only once during the given time period, no matter how many times it was seen in that interval, the observation was considered a sign system. If, instead of counting, the quality or frequency of observed behavior was to be estimated at the end of the interval of observation and

recorded on a continuum, it was probably a rating scale. The rating scale usually took the form of a scaled continuum which included perhaps five labelled points with its extremes signifying opposing conditions. For instance, one end would stand for 'almost always' or 'strongly' with the other end 'seldom' or 'weakly.' Both category and sign systems employed counting in contrast to rating scales which required judgmental estimating. The term category system was sometimes used to include sign as well as category type recording instruments. Of the three types, rating systems were, by far, the most numerous, and sign systems, the least commonly used (Rosenshine, 1973).

Regarding differences in items, as recently as 1971, the literature indicated that rating forms required high inference or processing through observer judgment and that sign and category systems were low inference measures. However, the recent proliferation of observation systems has blurred this demarcation of high and low inference as a distinction between category (or sign) and rating forms. Formerly it was thought that only general or global items were used in rating scales: "warmth, overall effectiveness, clarity, enthusiasm (Rosenshine, 1973:133)." And only specific items appeared in counting systems: "teacher gives directions, teacher asks divergent question (Rosenshine, 1973:133)." However, inspection and comparison of various observation systems revealed that specific behaviors were sometimes included in rating

scales thus producing lower inference measures (Rosenshine, 1973). It appeared that any item from a sign or category system could be used in a rating scale requiring the rater to indicate an estimated degree or frequency with which the specific behavior occurred. High inference characteristics have been introduced into sign and category systems. It was found that the same items could be rated on a rating scale during every time interval or could be tallied as in a sign system in each time interval in which they occurred. Thus, the type of item as well as the recording method was able to affect the degree of inference required.

Soar (1972) used four different category and sign systems in each observation of kindergarten and first grade classrooms in Project Follow Through. Thus, he had both category and sign type measures of the same activity and could compare results of the different systems. The variations in observation systems resulted in some functional differences. When a sign system was used for recording a set of behaviors, each item scored 0 or 1 in each time period of five minutes or so. If the same events were recorded as a category system, a tally was marked for each occurrence. "Events that occur infrequently record more weight with a sign system because the range of frequencies possible is limited (Rosenshine, 1973:134)" to one per time period. Frequent events recorded proportionately less weight with a sign system. The sign system was thought to distort the data or to provide a less

accurate reflection of the actual events. In the study reported by Ragosta, Soar, Soar, and Stebbins (1971), both sign and category systems were used for recording each classroom observation. Soar (1972) reported in regard to that study of selected Follow Through programs:

Ordinarily, an item is tallied only once in an observation period, but it seemed possible that the high rate of pupil response, which is emphasized in some programs, might be seriously underrepresented. As a consequence, the procedure of tallying each three seconds (or each interaction) was followed, but the data were analyzed as though they had been collected by both procedures. Conventional sign-system recording (tallying an item only once during an observation period) produced at least as clear factor structure, stronger differentiation of programs, and higher correlations with pupil growth measures . . . (p. 236).

Although a somewhat distorted presentation of the data, the sign scale seemed to be a better predictor of student gain. On the basis of this and other studies, Rosenshine and Furst (1973) generalized that:

. . . some observation systems which distort reality appear to be more predictive of student achievement than the systems which more closely represent the actual events . . . [and that no] one set of items, method of scaling, or format [can currently be considered] inherently superior to another (p. 136).

When results from many measuring instruments are compared, specific instructional variables, or clusters of variables, may be discovered to have consistent relationships with pupil change (McNeil, 1973).

Of all the observation systems, Flanders' Interaction Analysis (IA) System (1965) probably became the best known

and most used, either in its original form or in some modified or parallel version. Notable features of Flanders' observation system were attention to verbal behavior only, the three-second time interval used in recording interaction sequences, and the matrix format for scoring the interactions observed.

### Summary

Most of the works discussed in this section were descriptive studies. Their improved objectivity, low inference, and scoring systems made possible observer reliability and facilitated communication of data among researchers. Soar (1970) said of the use of systematic observation:

The possibility that systematic observation is the measurement breakthrough which will permit the development of a science of effective teaching seems very real (p. 121).

However, although the measurement of classroom process was greatly advanced, these studies generally lacked an additional step of correlating different aspects of Dunkin and Biddle's (1974) four-element teaching paradigm to each other.

### Relevant Studies

#### Importance of Process-Product Studies

Stake, known for his work in educational evaluation, quoted Cronbach's statement about the purpose of evaluation:

Cronbach urged another step: 'a most generous inclusion of behavioral-science variables in order to examine the possible causes and effects of quality teaching. He proposed that the main objective for evaluation is to

uncover durable relationships--those appropriate for guiding future educational programs. To the traditional description of pupil achievement, we add the description of instruction and the description of relationships between them. Like the instructional researcher, the evaluator seeks generalizations about educational practices. Many curriculum project evaluators are adopting this definition of evaluation' (Stake, 1967:526).

Thus was stated a plea for process-product research as an approach to analyzing effective instruction. Despite the fact that researchers expressed the importance of examining the relationships between process and product in education, by 1970, Rosenshine indicated there were only approximately thirty-five (1970a) or forty (1970b) studies relating observed behaviors to outcome measures:

Compared to the large number of descriptive studies, there have been relatively few studies of the relationship between measures obtained by the use of observational systems and measures of class achievement adjusted for initial aptitude or ability (Rosenhine, 1970b: 293).

Some of the correlational findings related to the current study were reviewed.

#### Flanders' Work

Although Flanders' Interaction Analysis (IA) System of recording observations was mentioned as a notable example of the instruments developed during the systematic observation movement, the Flanders (1965) study went beyond description. His work rightly deserves classification as a process-product or correlational study because he showed the relationship of

the instructional process to pupil learning. He concluded in his classic monograph, "Teacher Influence, Pupil Aptitude, and Achievement" (1965) that: "More flexible teachers (those he found to be more effective) were the most indirect when goals were being clarified and when new content material was being introduced (p. 112)." These same teachers were the most direct "after goals had been clarified and when work was in progress (p. 112)." Students of less flexible teachers learned less. "All types of students learned more working with more flexible teachers (p. 113)."

Flanders (1965) explained the implications of his findings for classroom teachers:

An indirect approach . . . is a way of providing the teacher with the student's perception of the situation, regardless of whether these perceptions are correct or incorrect. Such an approach . . . provides the teacher with more information . . . (p. 115).

about the learner's status and ways of learning. Flanders (1965) described the contrasting situation:

A direct approach increases student compliance to teacher opinion and direction. It conditions students to seek the teacher's help and to check with the teacher more often to be sure they are on the right track (p. 115).

Carrying his implications even further, Flanders felt that higher standards would be achieved not 'directly' by demanding that students perform in certain ways, but 'indirectly' by asking the students about their ideas and encouraging them to take greater responsibility for self-direction and for facing the consequences of their own decisions.

The teachers whose students learned the most were characterized by greater flexibility. As a result, the data showed these teachers to be less alike. The indirect teachers were "capable of providing many different roles and they shifted their roles (Flanders, 1965:116)" to meet different teaching purposes and/or learner needs. The poorer teachers were more alike and more direct. They "could not shift style of interaction . . . had fewer ways of working with students, and could provide only a limited number of roles (Flanders, 1965:116)." Thus the data made it easy to identify the poorer teachers due to their greater similarity but tended to identify the more effective teachers less definitively due to their greater disparity. The variety of roles and approaches used by effective teachers, therefore, added complexity to the task of defining measures of good teaching. Flanders' overall conclusion that, under more indirect teachers, students learned more and had more favorable attitudes toward school has been validated by frequent replication (Soar, 1970). Flanders' work and its influence on other researchers seemed to mark the beginning of a more fruitful study of instruction.

#### Documentation of Implementation

In contrast to Flanders' (1965) observation instrument which would measure verbal behaviors applicable to virtually any subject or academic level, some observation systems measured behaviors specific to a particular program or curriculum. An instructional program could be credited with increasing the



effectiveness of teaching-learning only if it was established that the materials and methods of the program were actually being used. Classroom observation was a way of determining the degree of implementation of a program. Rosenshine (1973) discussed implementation particularly in regard to Bissell's (1971) report of the Stanford Research Institute study "Implementation of Planned Variation in Head Start":

This type of description is rare at the present time. The documentation of implementation appears useful to any curriculum study or experimental study in which different treatments are being administered. Within the context of the Planned Variation research, differences in the level of implementation appeared to be extremely important during the first year of study. . . . Without data on implementation, comparative data on outcomes seems meaningless (Rosenhine, 1973:127).

Among sixty-one classes in eight different Head Start models involved, the twenty highest implementation classes were selected for comparison of pupil changes in academic achievement, general cognitive development, and response styles (Bissell, 1971). Implementation studies such as Planned Variation examined which instructional process factors were favorably related to cognitive changes in pupils.

Another example of process-product studies which featured implementation measures was Siegal and Rosenshine's (1972) report in which eight items stressed in the teacher training and the teaching guides were selected for observation. In twenty-four classes using the Bereiter-Engelman DISTAR program, the teachers were determined to be high implementers on the basis of specific rating scales developed

for each of the eight instructional behaviors stressed. In two studies, "three of the eight behaviors yielded significant correlations with student gain (Rosenshine, 1973:128)." Comparison of data from high implementing classes with parallel measures from medium and low implementing classes in the same programs exhibited potential for identifying which behaviors were most effective for cognitive gains.

In a three-year study of Project Follow Through programs, Soar and his colleagues (1972) recorded observations in seventy kindergarten and first grade classrooms. For each of the seven programs in the study, eight implementers and two non-implementers were observed. Four different observation instruments were used to record each observation. One of the four, Ober's (1970) Reciprocal Category System which focused on both pupil and teacher verbal behaviors, was a category system based on Flanders' (1965) IA System. The other three were sign systems: Teacher Practices Observation Record (TPOR), Brown (1968), Florida Taxonomy of Cognitive Behaviors (Webb, 1970), and Florida Climate and Control System (Soar, 1966). The data from each of the four observation systems were reduced by factor analysis. A study of variance was then conducted to find out whether the factors discriminated across programs:

. . . [A further] analysis correlated the factor scores with measures of class mean residual gain. These techniques are useful to determine whether a large number of educational variables discriminate among classes and are correlated with student growth. . . .The variables correlated with student

growth across all programs appear to be particularly potent, general instructional variables which are relevant (in a correlational sense) to many types of programs (Rosenshine, 1973:129).

In a study mentioned above, Soar (1972) (Ragosta, 1971) also found that indicators of "occasional, tight coercive attempts by the teacher to restrain students yielded a strong (negative) correlation with student growth (Soar, 1972:247)."

Soar's results, like those of Flanders and his forerunners, indicated that classroom interaction among pupils and teacher was of critical importance to observe and record in order to compare with pupil achievement regardless of the particular content area or instructional materials. Observing both implementers and non-implementers of each program and using the same observational and outcome measures in classrooms of each program made possible comparisons of the educational effectiveness among programs. Furthermore, the variables which were characteristic within particular programs could then be identified and related to outcome measures.

#### Variety of Approach

The quality of flexibility, thought to be applicable to all teaching situations, had implications in specific content areas when a teacher varied instructional approaches to fit different learners or learning tasks. Empirical evidence seemed to validate the use of a variety of instructional approaches in the following examples.

In reviewing process-product studies, Rosenshine and Furst (1971b) found ten categories of independent variables they thought strong enough to merit further research. One of the categories they defined as variability including various levels of discourse, various levels of tasks, and a variety of materials and techniques.

Gage's (1965) studies in teacher effectiveness indicated that good teaching outcomes were associated with teachers who were judged to have the qualities of warmth, cognitive organization, orderliness, indirectness, and ability to solve instructional problems. Perhaps the last mentioned could be interpreted as similar to flexibility or variety of approach.

Furst (1967) analyzed classroom data according to ratios of cognitive levels of interactions in the classroom--factual, inferential, and evaluative levels--and a question-answer-praise pattern versus a question-answer-criticism pattern. She found the most effective teachers were better by far in the use of variety of cognitive processes. She suggested:

These two types of behavior may be related: teachers who tend to behave in ways described as producing supportive climates also tend to use multiple cognitive levels when they deal with subject matter. There also seems to be some indication that these successful teachers tend to be more flexible . . . and vary their affective and cognitive behaviors more than do those who are less successful (Furst, 1971:178).

In reviewing Furst's work, Rosenshine (1971a) reported that the results of her study "suggest that the most effective

teachers exhibited greater variety in their use of questions (p. 86)."

Torrance and Parent's (1966) impressive School Mathematics Study Group (SMSG) report, "Characteristics of Mathematics Teachers that Affect Student Learning," attempted both to measure classroom interaction by observations and pupil questionnaires and to probe the thinking characteristics of teachers by means of teaching logs and self reports. "Gains in achievement and student aptitude in appropriate regression equations [were used] as criteria of teacher effectiveness (p. 2)." The analysis of teaching reports showed that the more effective teachers used proportionately more of the three higher mental operations--in Guilford's terms: convergent production, divergent production, and evaluation--and produced a greater variety of alternative ways of presenting mathematical concepts. The successful teachers ascribed the cause of their most successful or least successful lessons to teacher behavior. Conversely, the least effective teachers reported using greater proportions of the lower levels of thinking--cognition and memory, produced fewer alternative lesson approaches, attributed success in lessons to instructional materials, and lack of success to the learners and learning situation. These researchers suggested that teacher flexibility shown by the ability to use different techniques with different learners may indicate teacher awareness of learner differences and conscious attempts

to meet their needs (Torrance, 1966). Similar to Furst's (1971) findings, Torrance's indicate that use of a variety of higher cognitive levels may be associated with flexibility in teachers whose students make more gains. Torrance's successful teachers felt success of lessons was dependent on teacher behavior which seems closely akin to one of the qualities associated with effective teachers by Gage (1965): ability to solve instructional problems.

Some major findings in the field of reading were mentioned for their importance to this study even though they may not have been strictly process-product studies. Chall's (1967) voluminous report of the research which compared approaches in the teaching of beginning reading, her investigation of various reading programs, and her observation of reading classes concluded that there seemed to be no one approach that was best for all learners. Although there was some indication that beginning readers did better, at least in the primary grades, if they were taught by a code-emphasis rather than a meaning-emphasis method. However, no one of the code-emphasis methods appeared to be superior for all children. Chall conceded that some good teachers and some specific students profited more from other kinds of approaches.

For the United States Office of Education Cooperative Research Program in primary reading instruction (Bond, 1966; Bond and Dykstra, 1967; Dykstra, 1968), twenty-seven independent projects nationwide were selected on the basis of their

individual designs and their comparisons of varying methods of beginning reading instruction. All the projects collected their own data using the same measuring instruments and sent their data to a processing center at the University of Minnesota. Consequently, the effectiveness of a program could be compared to other programs or to the total data, and the accumulated information from all the programs could be treated as one massive study. Similar to Chall's conclusion, Bond (1966) found "no one approach so distinctly better in all situations and respects than the others that it should be considered the one best method nor to be used exclusively (Bond, 1966:8)." This reading research has also shown more variations among teachers using the same method than variations among methods which implied the importance of the teacher over all other variables in primary reading instruction. In both the Chall and Bond reports, the implication again seemed to be that the teacher had to know alternative strategies and be flexible in employing each when it was the most appropriate one for the learner.

Despite the favorable findings on the positive effects of teacher indirectness or flexibility or variability in the foregoing studies, there have also been some less favorable reports. Powell (1968) compared teaching behaviors with the achievement of 168 third graders who had been with the same teacher during their first three years of school. Similar process-product data were collected again the next year when

these children were fourth graders. The nine primary teachers of these pupils and the seventeen fourth grade teachers to whom these children were assigned were classified as direct or indirect on the basis of IA variables. The children assigned to indirect primary teachers for three years showed significantly greater gains in arithmetic achievement but not in reading achievement. However, by the end of their fourth year, even pupils who had been with only direct or only indirect teachers for all four years exhibited no significant differences in achievement. According to this result, differences in achievement were not maintained.

A process-product study by Soar (1966, 1967, 1970) dealt with grades three through six in four elementary schools. The process measures used in observing the fifty-five classes were Flanders' IA, Fowler's Hostility-Affection Schedule (1962), and part of Medley and Mitzel's OScAR (1958, 1959). There were thirty-nine items to observe. These formed nine factors of teacher behavior which were correlated with pupil cognitive gains in vocabulary, reading, arithmetic concepts, and arithmetic problems as measured by the Iowa Tests of Educational Development (ITED). Although four of the nine teacher factors correlated significantly with some pupil measure, the teacher factor which had a strong component of I/D ratios--indicating indirectness--did not show significant correlation.



Using the same data, Soar did another analysis. From within each of the two opposite types of teacher groups--the most warm and supportive and the most cold and critical--the direct and indirect teachers were identified. This resulted in the formation of four groups composed of the teachers at the four extremes of these two teacher dimensions. At each of the four grade levels, one teacher was chosen for each of the four extremes of the two dimensions: direct-high hostile, direct-low hostile, indirect-high hostile, indirect-low hostile. When Soar compared each teacher type with the vocabulary and reading scores of pupils in those classrooms, each produced a significant correlation indicating that indirect teachers were more effective than direct teachers. Thus, Soar's data, when analyzed in different ways, appeared to produce different results.

Rosenshine (1973) felt that many observational systems could be used in classrooms of various content areas and levels of education. Soar (1972) found factors derived from data obtained using four different observation instruments in each of seventy classrooms across seven different Follow Through programs. Soar compared these classroom factors with the class means of pupil gains and found that there were a number of general variables which correlated with pupil gains across grade levels and in various types of instructional programs. Soar enumerated the variables he found associated with greater pupil gains as warm, accepting classroom atmosphere; pupil freedom, initiation, and self-direction; and

teacher control and prescription which increase simple learning, but only up to a certain point (Soar, 1972).

In the USOE Longitudinal Study of Educational Practices--Project LONGSTEP--conducted by the American Institutes for Research (AIR) (Coles et al., 1976; USOE, 1976), an attempt was made to determine the educational outcomes in schools and programs employing high degrees of individualization or innovative practices. These investigations involved thirteen school districts in nine states, including 30,000 students and 1500 teachers in 80 schools during the three-year period 1970-1973. The unexpected findings indicated that neither intensive innovation nor a high degree of individualization was able to induce "substantial yearly gains in student achievement (Coles, 1976:19)." Student and teacher questionnaires were used to assess presage variables and attitudes. An Educational Experience Analysis Guide was developed to determine similar and variant characteristics of programs. On the basis of specific, observable characteristics, each school program was located on a continuum from traditional to innovative. Students were classified not by the program in which they were enrolled, but according to their educational experiences which divided them into over two hundred groups to indicate amount of innovation and degree of individualization. A classroom observation instrument was designed to record:

. . . physical environment, study arrangements,  
and access to resources, as well as teacher and

student activities such as degree of grouping, focus of activities, and use of materials (USOE, 1976:2).

Site visits of five to eight days were conducted three times a year for three succeeding years. During site visits, data were collected from school records and principal and teacher interviews, as well as classroom observations to determine the educational experience classifications. In this project, a student was assumed to be exposed to a specific process variable only if the data documented implementation of the variable in that student's classroom. Selected items of the Educational Experience Analysis Guide were used to arrive at an index of the level of innovation which included a factor called the degree of individualization. These two measures were considered the instructional variables which were then related to student achievement test performance. Results showed that:

The single most important and well documented finding was the lack of either substantial or consistent association between student achievement and overall level of innovation across grades (USOE, 1976:4).

Thus, according to the results of this study, educational innovation and individualization by themselves should not be expected to improve educational attainments. If it could be assumed that degree of innovation and individualization implemented in a classroom were indicative of a teacher's ability to employ variability and flexibility, the LONGSTEP findings seemed to conflict with results of teacher flexibility as measured by Flanders' IA.

The kinds of instructional activities conducive to progress for some kinds of learners were less effective for some others. Bennett (1976) reported on several studies which illustrated this point. The relationship of language, spelling, and arithmetic test scores of a random sample of third graders from structured and unstructured classrooms showed that compulsive pupils did better in structured classes than less compulsive pupils in the same or in unstructured situations. Compulsive pupils showed no differences in unstructured settings. Anxious pupils did about as well as the average student in structured classes. However, anxious pupils in unstructured settings apparently felt threatened and achieved significantly less than low anxious pupils in unstructured classes (Grimes and Allinsmith, 1961).

In fifty-four classrooms of grades three through six, Soar (1968) used four observation instruments as process measures and four product measures: vocabulary and reading subtests of the Iowa Tests of Basic Skills, the Toy Dog Unusual Uses Test from the Minnesota Tests of Creating Thinking, and the Children's Manifest Anxiety Scale. This research included a comparison of results for high anxious and low anxious pupils. Both types of pupils learned more with more indirect teaching. Low anxious pupils benefited the most from less teacher control. Soar considered his findings in this study as hypotheses for further research.

Another kind of learner difference was studied by Amidon and Flanders (1966). They found in geometry classes that "dependent-prone" pupils were more sensitive to type of teacher control. Such children varied in amount of achievement when they were exposed to different types of teacher behavior. The less "dependent-prone" pupils showed less reaction to varying teacher styles in terms of achievement levels.

Bennett cited two studies which indicated that differences in learner ability levels interacted with teaching styles. Schantz (1963) found that high ability students exhibited greater gain under indirect than under direct teaching while there was no difference in the effect of teaching style for low ability students. Calvin, Hoffman, and Harden (1957) also found that permissive teaching was an advantage for high IQ students but a handicap for pupils with average IQ.

Mills (1956), in teaching word recognition skills to second and third grade readers who were six months below expected reading levels, compared four approaches: kinesthetic, phonic, visual, and a combination of all three. Results were determined on the basis of ability to learn ten words chosen by their high frequency use in basal readers. Pupils of IQ 80 and below made the most progress with the kinesthetic approach, but not significantly better than with the visual or combination approaches. The least effective overall approach, the phonic, resulted in significantly smaller results for

these low ability learners. Pupils of IQ 85-100 responded best to the visual and combination approaches, only slightly less well to the phonic, and least well to the kinesthetic approach. More capable children, IQ 105-120, did about equally well in all approaches. Thus, differences in learner ability levels were strong determinants among teacher effectiveness measures.

Contrary to research which seemed to show that process factors such as teacher flexibility were effective across age or grade levels, there have been some results showing that learners of different ages responded differently. Bennett (1976) cited work of Powell, Flanders, and Wallen. The first showed that pupils whose achievement gains correlated with the degree of directness of their teachers at the end of third grade did not show similar correlations with their similarly classified teachers' styles at the end of fourth grade. Differences among learners of different grade levels were shown in Flanders' research by positive correlation of upper grade students' cognitive growth with teacher sustained acceptance of student ideas contrasting with a negative correlation for the same interaction in a second grade sample. Wallen found similar contrasts between the relationships in first and third grade samples.

In nationwide studies of First Grade Reading Instruction, Dykstra (1967) found that:

. . . the least mature pupils achieved better in a Basal program than in a Language Experience approach, while more capable students . . . [in auditory discrimination and letter knowledge] profited more from a Language Experience approach (p. 11).

Dykstra's (1967) findings also showed that, in general, girls were superior to boys in readiness and achievement measures of first and second grades; girls tended to be better in all programs tested; and no differences in reading achievement were found between boys and girls when the achievement scores were adjusted for differences in readiness. However, Coles (1976) attempted to explain some of his results:

The undoubtedly dramatic growth in achievement demonstrated by a number of students particularly in Project LONGSTEP certainly suggests that some near-optimal match of student and educational approach may have been one of the reasons for the gains of these students (p. 28).

Thus, teacher behavior was shown to have different effects depending on learner differences such as compulsivity, anxiety, intellectual ability, dependent-proneness, or age levels and the interaction between teacher behavior and pupil idiosyncracies.

The Beginning Teacher Evaluation Study: The Effects of Teaching Performances on Student Learning (BTES) (McDonald, 1976a,b,c,d) also included some results which related to the differences among learners of different grade levels. BTES was a research project sponsored by the National Institute of Education and conducted by the Educational Testing Service, Princeton, New Jersey, for the California State Commission for

Teacher Preparation and Licensing. Its aim was to learn which teaching behaviors significantly affect what and how pupils learn. This study proposed:

. . . to (1) develop an assessment system for measuring teacher and pupil behaviors and other factors which could influence each of them and their interrelationships and (2) generate hypotheses about the interrelationships among teacher and pupil behaviors and related factors (McDonald, 1976a:abstract).

Data were collected on the learning process by systematic observation in classrooms of forty-one experienced teachers of second grade and fifty-four of fifth grade in forty-three schools in eight school districts in California. Pupil achievement in reading and math were used as measures of instructional product: reading scores for decoding, comprehension, and applications, and math scores for computation, concepts, and applications. In addition, much information was amassed on both the pupil and teacher presage factors: student attitudes, aptitudes, cognitive style, and expectations; also teacher knowledge and aptitude factors including cognitive style. Data were also gathered about the educational context: administrative organization, climate, and responsibility of the teacher. All were then related to variations in teaching performances. The teacher served as the unit of analysis and each analysis included data for all the pupils who had both spring and fall test scores. The classroom observation instruments used were a narrative behavioral recording system, Anecdotal Process for Promoting



the Learning Experience (APPLE), which focused on pupil activities and pupil-teacher interaction (Lambert, 1976a,b), and a categorical system, Reading and Mathematics Observation System (RAMOS) (Calfee, 1976a,b), to record teacher behaviors during reading and mathematics instruction (McDonald, 1976a,b). In addition, teacher reports of their plans and goals for the year in reading and mathematics and structured diaries of their daily planning and teacher activities for two designated weeks during the school year were obtained (McDonald, 1976b). A major finding of the BTES was that:

. . . no single skill or teaching performance was found to be equally or comparably effective in both [second and fifth] grades or in both [mathematics and reading] subjects (McDonald, 1976d:48).

If this finding were confirmed in further research, it would imply that:

. . . the goals of training teachers for the primary or the intermediate grades and the criteria for evaluating their competence will necessarily be different (McDonald, 1976d:48).

The BTES led to the conclusion that:

. . . there are no single teaching-performance variables correlating so significantly with children's learning that they should be considered critical for effective teaching. . . it is different patterns and structures of teaching acts that influence changes in learning rather than single, omni-effective teaching performances (McDonald, 1976d:49).

There emerged two hypotheses from the BTES data:

. . . a pattern of teaching practices is more likely to be related to learning than is a single practice [and] effective teaching patterns will differ by subject matter and by grade level (McDonald, 1976d:55).

The differences in effective teaching behaviors were probably due to the differences in what is to be learned at the two grade levels as explained by McDonald (1976d):

At the second grade discrete responses are being acquired and linked together. Therefore continuous instruction for individuals and monitoring of the acquisition process are probably needed. At the fifth grade cognitive processes to be used with varied content are being learned. Teaching strategies which stimulate comprehension processes are probably required (p. 6).

As the BTES conclusions and other research have pointed out, teaching behaviors may need to be adjusted to differences among learning tasks as well as to the differences among learners discussed above. Not only were the learners of different grade levels at different stages of maturity, but also the learning tasks shifted from acquiring and linking discrete skills and concepts to the application of those skills and concepts in the more abstract and complex, higher level cognitive processes of comprehension (McDonald, 1976b).

Soar's (1966, 1968) work with over fifty classes of advantaged children in grades three through six indicated that different degrees of pupil freedom were optimal for different tasks. Relatively great pupil freedom and only moderate levels of teacher control resulted in the greatest growth in complex learning tasks such as vocabulary learning or acquiring new math concepts. Greater teacher control was associated with more gain in simple concrete learning such as reading, arithmetic facts, or spelling words.

Similarly, results of Soar's (1972) Follow Through investigations suggested that:

. . . increased amounts of drill may be functional for simple concrete growth and that an optimal balance between pupil initiation and drill is a condition for complex-abstract growth to take place. But the aspect of the relationship which appears strongest is the indication that greater than optimum amounts of drill are strongly destructive for complex-abstract growth (p. 254).

In discussing four observational studies, Rosenshine (1971a) found them difficult to compare because of their widely divergent observation instruments. Nevertheless, he pointed out that the results of each of the studies emphasized that "patterns of behaviors are more important than single behaviors (p. 86)," and that these behavior patterns were optimally effective somewhere between their extremes. Moderate use of questions and small amounts of drill (Soar, 1966; Solomon, 1963) and the use of a variety of question types (Furst, 1967; Thompson and Bowers, 1968) were examples of the variations in teacher behavior patterns they found most effective for pupil achievement.

Thus a teacher who could vary behavior from highly controlled and focused for teaching simple learning tasks to an indirect teaching set which allowed much more pupil initiation and freedom for promoting abstract learning and divergent thinking seemed to be most effective for overall pupil gains. A teacher who used various instructional approaches increased the possibilities for optimal learning.

In summary, teacher variability and flexibility appeared to have critical potential for increasing pupil achievement. Classroom activities which indicated patterns of shifting from one approach to another for different kinds of learning tasks and the flexibility to vary the learning approach for different kinds of learners appeared to have more potential for effecting learning change than a high degree of any one instructional behavior. However, instructional research must serve the function of correlating both general instructional variables and program-specific variables with pupil gain to determine the relationships which have the greatest value for improving instruction (Rosenshine and Furst, 1973). Variables which correlated with student gains across programs were considered the most important to develop and implement (Soar, 1972).

### Teaching of Skills

Some specific aspects of teaching have been studied to determine their effects on learners. Several of these instructional factors related to the present study: the sequence in which the specific skills of a discipline are acquired, the diagnosis of learning difficulties, the direct teaching and reinforcing of specific skills, the amount of focus and control provided by the teacher, and the amount of time spent on the learning tasks.

In the sequence of the intertwining language arts skills, listening and speaking preceded reading and writing.

In Dykstra's (1968) report on the First Grade Reading studies, he concluded in part that knowledge of phonics helped learners to recognize words more readily and to spell better, and that practice of writing skills improved progress in primary reading. Studies of Soar (1972) and Conners and Eisenberg (1966) demonstrated the importance of oral communication between pupil and teacher. Soar (1972) inferred from his data that the disadvantaged children in the Follow Through programs "profit from extensive experience with the simple encoding and decoding of behavior and experience into language (p. 250)." He also found evidence that exposure to a model of teacher talk followed by "sustained, self-initiated pupil talk (p. 251)" seemed to be the most valuable for promoting abstract learning. From this information, Soar inferred that there existed a "need for a model for pupils before it is functional for them to be involved in extensive talk (p. 250)." Conners and Eisenberg (1966) found that the total number of 'communication episodes,' especially those with cognitive content, correlated significantly with intellectual growth-related activities. Harris (1966, 1968) found that a variable called 'total interchanges' correlated significantly with pupil measures of word reading, word study, and spelling on the Stanford Achievement Test at the end of first grade, although not with measures obtained by using the Metropolitan Achievement Test at the end of second grade. Among the behaviors Fortune (1966) found associated with the most effective Operation Headstart teacher trainees were allowing pupils to

handle and try out real objects and teacher modeling of verbal patterns for pupils to repeat. It was inferred from the above examples that concept formation and the development of oral language patterns and phonic skills occurred early in the sequence of language learning and formed a basis for subsequent successful reading achievement.

Gage, Rosenshine, and others have studied specific aspects of teaching they thought critical for effectiveness. One of these factors was the diagnosis of learning difficulties (Rosenhine, 1973). The most effective teachers in the SMSG (Torrance, 1966) used more of the evaluative behaviors categorized as trouble shooting, diagnostic evaluation, and hypothesis making and testing than they did behaviors classed as negative evaluation or positive evaluation. Torrance's study indicated that it was not necessary for every child to be taught every lesson in the basic text. Teachers who exhibited awareness of differences among learners and diagnosed pupil needs were enabled to teach specific skills selectively and thereby meet varying needs. Diagnosing learner difficulties served to pinpoint what needed to be taught.

In regard to reading, both Chall (1967) and Bond (1966) emphasized that "no matter what the underlying method is, word-study skills need to be emphasized and taught systematically (Bond, 1966:9)." Chall (1967), in recommending a code-emphasis method of teaching reading, specified the advantage of the direct teaching of specific skills of letter-sound

correspondence and the value of writing, tracing, or typing as adjunct skills in learning to read. Dykstra (1968) agreed that the method of teaching phonics was not as important "as the fact that direct attention is given to helping pupils learn sound-symbol relationships (p. 7)" to provide ability in word recognition and spelling achievement in primary grades. Dykstra (1968) also pointed out that: Direct instruction in comprehension is apparently essential even in beginning materials (p. 10)" since pupil achievement in word recognition and spelling was not automatically related to achievement in comprehension. In addition: "Generally direct vocabulary instruction results in greater progress than does incidental instruction or wide reading (Robinson, 1971:408)." As the above studies indicated, most kinds of skills in reading had to be taught directly and were not usually learned incidentally nor as a result of acquiring other reading skills. Rosenshine and Furst (1973) may have been referring to the direct teaching of specific skills when they suggested that instructional research in connection with curriculum and materials should include "monitoring of opportunity to learn the criterion material (p. 130)." Dykstra (1968) concluded in part that direct instruction in comprehension skills was essential.

Practice of the specific skills that were being taught showed positive correlations for first graders (Wallen, 1966). When teachers gave assent, but not necessarily strong praise,

and asked frequent questions, pupils apparently experienced a practice and reinforcement effect and achieved more. Behaviors which marked teachers who were more effective in teaching a specific content material included introducing the lesson to provide instructional set, employing review and repetition techniques during the lesson, reinforcing student responses, and integrating pupil answers into the lesson (Fortune, 1967). Similar to Wallen's study, Fortune's indicated that frequent practice and reinforcement encouraged favorable outcomes. Conners and Eisenberg (1966) reported that 'communication episodes' which had cognitive content were more frequent with teachers of high-achieving classes and were less frequent with teachers of low-achieving classes. Interchanges which dealt with nonintellectual topics such as care of materials apparently distracted attention from cognitive tasks. Two other studies also qualified the effects associated with practice. Furst (1967) found that certain kinds of questions and moderate use of 'teaching cycles' were used by high-achieving teachers. Soar (1972) determined that too narrow a focus and too much drill was destructive for complex-abstract learning although it was productive for easily measured simple-concrete learning. So, although review reinforced by teacher approval was necessary for increased pupil achievement, under certain circumstances, structured practice was found to be counter-productive.



In addition to the importance of practice and reinforcement, the instructional set or focus provided by the teacher has shown promise of being a critical factor. The successful teachers were found to provide a presentation designed to focus the mental set for pupils before questioning them (Soar, 1966; Furst, 1967) and to follow up by probing for further explanation by pupils of their responses (Soar, 1966; Spaulding, 1965; Fortune, 1967). Not merely the frequency but also the types and objectives of questioning affected pupil achievement. Questioning associated with the more effective teachers included more questions focused on academic content rather than on personal interests or belongings (Spaulding, 1965; Conners and Eisenberg, 1966) and used a combination of question types requiring explanation, clarification, interpretation, or judgment by the student (Furst, 1967; Solomon, 1963; Fortune, 1967; Soar, 1966). Both the teacher questioning techniques and the pupil responses which received teacher approval influenced achievement most when they were focused directly on the cognitive goals of the teacher. Rosenshine (1971a) explained what correlational studies have demonstrated:

After the primary grades, single cognitive behaviors are not significant correlates. Rather, the over-all pattern of behaviors is more important. Such a pattern includes the use of a variety of questions, moderate amounts of structure, lesser amounts of drill, and frequent requests for the pupil to elaborate his answer (p. 93).

In addition to focusing on intellectual problems by structuring mental set, by using various question types, and by probing for expansion of pupil answers, the effective teachers varied amounts of control and pupil freedom to match the instructional purpose (Soar, 1970, 1972, 1976). In Soar's Follow Through Classroom Process Measurement (1976), pupil learning was found to be related to the amount of teacher control and its opposite, pupil freedom. The relationships represented on a graph resulted in curves rather than straight lines. The curve representing simple-concrete learning--such as classifying according to a single attribute, or counting, or matching or naming shapes, letters, or numerals--showed the least learning of this type occurred toward the end of the continuum representing the greatest degree of pupil free choice. Slightly more learning than the minimum occurred at the extreme of pupil freedom, and the most achievement was gained at the opposite extreme, strong teacher control with almost no pupil self-direction. Thus simple-concrete learning increased most with direct teacher control of the activity. For skill activities, a somewhat similar curve was even more accentuated: small gains with pupil initiation and even greater amounts of cognitive growth after drill exercises structured by the teacher. In contrast, the amount of complex-abstract learning increased from the pupil-choice end of the continuum to a maximum of learning where there is some teacher control but more pupil freedom. The amount of complex-abstract learning declined somewhat with near absence of

teacher control and almost complete pupil freedom. In the other direction, complex-abstract learning also declined steadily as the proportion of teacher control increased.

The CRAFT Project (Harris, 1966, 1968) also examined the effect of teacher control on pupil learning. Harris used observations of forty-eight classrooms of first graders and thirty-eight second grade classrooms of the same pupils the following year. The observation instrument used was a special version of the OScAR modified for observation of reading instruction and named OScAR-R. In the first grade data (Harris, 1966), the variable called 'control,' which included items such as criticism of students, was related to pupil scores in word reading, paragraph reading, vocabulary, and word study on the Stanford Achievement Test. Control correlated negatively to a significant degree with the measures of pupil achievement. In the second grade data (Harris, 1968), scores on the Metropolitan Achievement Test were used, and control again correlated negatively with achievement but not to a significant degree. Thus the amount of focus and control exerted by the teacher seemed to affect pupil attainment, and the type of learning to be accomplished determined the optimal combination of teacher structure and pupil freedom to employ.

Another process factor which was found to influence learning outcomes was called on-task time: the amount of instructional time spent per day in teaching and learning the skills. Travers (1971) wrote:

The most important of the often neglected variables is time. One independent variable may appear to be more effective than another when actually the differences in time spent under each teaching condition may be the significant factor of variation (p. 30).

Data on the amount of instructional time spent on a subject each day was found to relate positively to pupil achievement as reported in Project LONGSTEP (USOE, 1976):

. . . Students who exhibited unusually large gains in reading and language achievement during two consecutive school years (grade three) had been exposed to much more class time on these subjects as second graders than were students with a notable lack of growth.

. . . the findings suggest that increasing the amount of class time per day for language arts, may be a worthwhile strategy to improve student performance and that the greatest pay-off may come from a concentration of such efforts in the early elementary grades (p. 4).

In fact, Coles (1976) concluded that increasing the amount of class time spent in language arts instruction per day, especially in early grades, gave evidence of helping to improve achievement even at later grades. Even though the pupils who had made dramatic gains in the second grade program with, as Coles (1976) noted,:

. . . a notably greater amount of class time per day on language arts . . . [and] spent considerably less time per day on language arts in grade three, the overachievers again demonstrated dramatic gains (p. 23).

Teachers in the BTES (Elias, 1976c) described their instructional activities during two specified weeks by recording information in Work Diaries. These teacher reports were compared with observational data and pupil achievement

measures. The structure and content of the diary reports included information about the amount of time teachers spent planning for and teaching reading and math skills. Although the amount of direct teaching time did not correlate with pupil gains in math skills, the instructional time did relate to second grade pupil gains in decoding skills.

The kinds of activities included in the language arts time period were critical as indicated by Harris' (1966, 1968) significant negative correlations of pupil gains in reading achievement with amount of time used for reading stories to pupils. The Harris and Serwer (1966) results showed that the factor most influential in increasing reading achievement for most first grade children was the time allotment: not time spent in class management, discussion, or art activities, but the amount of time spent in reading activities. The Harris and Morrison (1969) Final Report on the CRAFT project indicated that there were greater differences between class mean reading scores within each group using a reading method in common than there were differences between the means for the different methods. In general, increasing instructional time in each approach to reading instruction increased achievement levels. Similarly, Conners and Eisenberg (1966) reported negative results for reduction in on-task time caused by interactions which dealt with pupils' property and materials rather than with content. And McDonald (1976c) reported as a negative predictor for fifth grade reading achievement the

amount of time the teacher used giving procedural directions rather than content instruction. According to another phase of this same study, reports of teachers about time spent teaching specific skills appeared to be reflected in evidence of pupil skills in those areas (Elias, 1976c). However, this study also showed a result similar to Soar's findings regarding simple skills:

Greater amounts of time spent teaching the more specific decoding and vocabulary reading skills to fifth grade pupils was associated with poorer pupil reading performance on the more generic reading tasks, comprehension, application, and achievement (Elias, 1976c:326).

Thus, effectiveness in teaching the skills was influenced by the sequence in which the skills were taught, the diagnosis, direct teaching and reinforcement of specific skills, the degree of focus on content and control of instructional activities exerted by the teacher, and the amount of time per day spent on cognitive tasks. In addition, the amount of teaching-learning time spent on a subject per day was found to be of very great importance especially in the primary grades and even to have residual effects.

#### Classroom Interaction

Foregoing sections of this review have dealt with variety of approach in instruction and the diagnosis and teaching of specific skills. The primary concern of both sections was with adjusting intellectual aspects of instructional behavior to specific learning tasks and learner needs. Another facet

of classroom behavior which appeared in research reports dealt with affective variables involving learning climate, especially the personal interactions such as peer relations; teacher acceptance and support of pupil ideas, pupil decisions, and pupil self-evaluation; and teacher encouragement of pupils' critical thinking and creativity. Such affective variables, involving personal relationships, were more consistently related to achievement across grades and content areas than were cognitive variables which tended to be program-specific variables. Many of the factors discussed above in regard to teacher flexibility and variety of approach were found to affect the quality of interaction as well as the cognitive outcomes. The teachers who utilized the flexibility to cue diverse instructional approaches to the pupils' cognitive needs and abilities were usually at the same time engendering classroom climate and interactions which affected pupils' attitudes and responses to the learning situation. Studies to discover the correlation between general instructional variables and pupils' achievement gains is considered one of the four priority areas in instructional research (Rosenshine, 1973).

Among groups of climate variables that predicted learning better than others, Wahlberg and Anderson (1968), in studying "Classroom Climate and Individual Learning," found that 'synergism'--personal relations among class members--predicted learning better than 'syntality'--identification

with group goals. The same study indicated that measures of 'isomorphism'--the tendency for class members to be treated equally--predicted learning better than other variables tested (p. 418). Thus, pupils learned better when they felt accepted by others in their microcosm.

In the research he reviewed, Gage (1965) found that five global characteristics of teachers emerged as having definite relationship to effective teaching: "warmth, cognitive organization, orderliness, indirectness, and ability to solve instructional problems (p. 88)." The last two mentioned have been discussed above in regard to variety of instructional approach. Of Gage's five qualities, warmth and indirectness were most closely associated with emotional climate. Teacher warmth was described as 'acceptant' behavior, using criticism in small doses resulting in a higher proportion of acceptance than of rejection toward pupils, and a 'threat-free' climate where pupil self-expression and active pupil participation were engendered (Wood, 1970). In a warm climate, students readily initiated activities apparently without fear of rejection or negative criticism from teacher or peers.

Brown (1970) designed an observation system to measure the kinds of process factors found to flourish in a warm learning climate. Called the Teacher Practices Observation Record (TPOR), this sign system focuses on elements of Deweyan experimentalism such as pupil-centeredness, active pupil



participation, allowance for pupil initiative, teacher acceptance and extension of pupil ideas, and encouragement of intrinsic motivation. These process factors exemplified behaviors associated with a warm learning climate (Wood, 1970). Brown (1970) indicated the general applicability of the factors to be observed by reporting that items of the TPOR have been used as behavioral objectives in kindergarten through college level classes in any content area for either teacher assessment or on-going school evaluation. The effects of teacher criticism, the antithesis of warmth, were discussed by Soar (1968) in "Optimum Teacher-Pupil Interaction for Pupil Growth." Complex-abstract learning was optimal with the least teacher criticism. Both simple-concrete learning and divergent thinking increased with small amounts of criticism. Then, as criticism increased beyond an optimal small amount, the learning of simple material and complex learning both decreased, complex learning suffering the most from much criticism. Soar (1968) stated that all three kinds of cognitive behavior necessitated the teacher's being warm and supportive of pupils but more direct in teaching style for some objectives than for others.

The two dimensions, warmth and directness, have often been pooled in the term permissiveness. To Soar, it appeared that these were distinctive and exerted different effects in the learning situation. Warmth was described as an aspect of emotional climate. Directness was a factor having to do with

degree of teacher control. Regarding teacher control, Soar (1966) said further that indirect control by the teacher during the academic year was found to result in more growth during the summer than extreme use of direct control.

An exceptionally effective means of expressing teacher acceptance, support, or positive reinforcement of pupil initiative or pupil response has been teacher use of pupil ideas. Furst (1971) explains:

Of the approximately fifty studies done as of 1970 with systematic observational techniques in the area of teaching effectiveness, most have concerned themselves with overall affective (climate) dimensions of teaching behaviors. There is a convincing amount of data which more than implies that teachers who generally use student ideas for some periods of time and those who build on student ideas are teachers whose students have higher than average achievement on tests of information at different grade levels. These students often also have positive attitudes toward school and subject matter under study as well as lower levels of anxiety and more positive self-concepts. These results are fairly consistent across grade levels in both short-term studies and studies over longer periods of time (p. 177).

Flanders (1965) pointed out that the use of pupil ideas by the teacher was a use of indirectness by the teacher:

Some critics of the public schools have advocated that teachers 'get tough,' tell students what to do, and demand higher standards. Our data show that higher standards can be achieved not by telling students what to do in some misguided 'get tough' policy, but by asking questions and then using student ideas and perceptions and reactions to build toward greater self-direction, responsibility, and understanding. If 'getting tough' means helping students face the consequences of their own ideas and opinions, then our indirect teachers are much tougher (p. 116).

Flanders and Simon (1969), reporting on studies of teacher effectiveness, found eight studies which showed a positive relationship between pupil achievement and the percentage of teacher talk utilizing pupil ideas. One of these studies showed in addition that pupils in such classrooms asked a type of thought-provoking question that occurred very seldom in most classrooms (Johns, 1966). Flanders and Simon (1969) also mentioned three studies that did not support a relationship between teacher use of pupil ideas and pupil achievement. Bennett (1976) stated that since the introduction of the systematic observation movement with its more objective measures, Flanders' indirect teaching behaviors--accepts feeling, praises and encourages, uses pupil ideas--have shown steady consistency.

Pupil participation in decision making, planning, responsibility for self-evaluation and intrinsic motivation seemed generally to increase in classrooms with indirect teachers. On the basis of his IA studies, Flanders (1965) drew some implications for classroom teachers:

An indirect approach will stimulate verbal participation by students . . . [and] it often results in the students developing more responsibility for diagnosing their difficulties and suggesting a plan of action (p. 115).

The data from Project LONGSTEP (Coles, 1976) seemed to contradict Flanders' conclusions. LONGSTEP attempted to determine a measure or index of the amount of innovation and the degree of individualization pupils experienced. These indices were based on such factors as decision making,

instructional pace, and use of performance agreements which, when taken together, seemed to have much in common with the factors noted by Flanders in the classrooms having indirect teachers. In spite of group differences in achievement; no consistent overall relationship was found between innovative intensity, as shown in LONGSTEP's measure of innovation, and posttest reading performance, or between individualization emphasis, as shown by LONGSTEP's measure of individualization, and posttest reading performance (Coles, 1976). In fact, Coles (1976) pointed out:

. . . in the only series of analyses in which growth in achievement was related to overall innovation and individualization to a substantial degree (during third grade), the impact of Level of Innovation (or Degree of Individualization) was negative--. . . the greatest growth occurred in programs with a more moderate emphasis on innovation (p. 21).

Analysis of data about individual students showed that consistent overachievers tended to be members of programs with less innovation and individualization. However, the degree of individualization did not show consistent positive correlation with pupil achievement in Project LONGSTEP.

Among McDonald's (1976b) major conclusions in the Beginning Teacher Evaluation Study (BTES), teacher behaviors were determined to be responsible for about a third to a half of pupil achievement gains. "Results [of this study] indicate a significant and consistent effect of teaching performances on student learning (p. 4)." The BTES Phase II report (McDonald, 1976a) stated it was important to include

the description of student behavior since student activities are the critical events which affect student learning (p. 224). McDonald seemed to agree with Morsh's (1956) report that pupil behaviors were even better indicators of teacher effectiveness than were teacher behaviors. Of teachers, McDonald said ". . . differences in effective performances are probably related to the differences in what is to be taught (1976c:319)" in different subjects and grade levels.

Soar (1964) and Mitzel (1960) indicated that they thought neither teacher behavior nor pupil behavior can be analyzed exclusive of the other:

In considering both teacher behavior and student behavior as process criteria it becomes clear that neither of them should be studied in isolation from the other. The interaction between them appears to be the dominant aspect of the whole process of learning (Mitzel, 1960:1484).

In a similar vein, Coles (1976) said:

The undoubtedly dramatic growth in achievement demonstrated by a number of students particularly in Project LONGSTEP certainly suggests that some near-optimal match of student and educational approach may have been one of the reasons for the gains of these students (p. 28).

One of the several papers in connection with the BTES described the APPLE Observation System which was used for documenting classroom activities. From the study of APPLE records, teacher behavior in classrooms where children made both high and low achievement gains for the year were compared. It was found by Lambert (1976c) that:

. . . not only do more effective teachers engage in certain teaching activities more frequently but they also employ a wider variety of activities. They conduct instruction by actively interacting with children, checking the work, asking questions, giving instructions and checking for understanding with question-and-answer sessions (p. 323).

The cognitive effects of pupil-teacher interaction were shown graphically by Soar. Teacher control and pupil creativity or divergent thinking appeared to bear an almost linear relationship (Soar, 1968). Pupils showed the least creativity under the most structured teacher control, creativity increasing as teacher control diminished. Higher level reading skills such as inference, integration of new information read with the reader's previous knowledge and experience, predicting and evaluating involved aspects of creativity. Studies regarding creativity probably were also applicable to reading of abstract material such as modern poetry, according to Soar (1967):

It seems reasonable to expect that the development of complex reading skills should be related to the development of creative processes, and therefore, should be influenced by the same environmental factors (p. 245).

### Summary

Classroom interaction consisted not only of teacher behavior nor only of pupil behavior but depended on the interaction of both. How well teachers and pupils interacted in classroom endeavors determined the learning climate. The quality of classroom interaction was found to have strong influence on pupils' cognitive progress.

### Summary

A review of the classic observational studies of teacher effectiveness emphasized the difficulties inherent in attempts to measure educational processes. The design of teaching models helped researchers to classify and relate different aspects of instruction. The development of the systematic observation measures and methods increased the potential for reliability in describing classroom processes.

In determining the effectiveness of teaching techniques, a number of studies went beyond the low inference, objective description of classroom processes to the correlation of process measures with instructional measures. Correlational studies have used cognitive, program-specific variables as well as affective, general variables which were measurable across classrooms involving different content areas, age groups, and instructional approaches. To establish the efficacy of an educational program, it was necessary to ascertain the degree to which the program was implemented in the classroom.

A review of relevant correlational studies suggested cognitive and affective factors which appeared to be indicators of effective instruction. Of particular interest in the current study were those factors which related to three general aspects of elementary reading instruction: variety of instructional approach, systematic teaching of skills, and quality of personal interaction in the classroom.

## Chapter 3

### DESIGN AND PROCEDURES

This study was designed to compare a systematic description of classroom processes with a standardized measure of reading achievement as a means of evaluating the Elementary Reading Improvement Program. Important aspects of the design and procedures of this study were (1) a description of the design of the study, (2) the development of an observation checklist, (3) the selection of the stratified sample, (4) the collection of observation data, (5) the administration of achievement tests, and (6) the treatment of the data. This chapter closes with a brief statement of the research design.

#### Design of the Study

In order to determine the effectiveness of the ERIP, the teaching-learning process in classrooms was compared with the resultant change in reading achievement levels made by the pupils in those respective classrooms. An observation checklist was developed for the specific purpose of measuring the classroom instructional process. A large part of this study was therefore concerned with the formulation and modification of the observational instrument. Finally, the instrument was used to record observations of classroom



activities by which the degree of teacher implementation of the reading program was established.

On the basis of an initial observation conducted in every classroom which met the criteria of this study, the classrooms were rank-ordered according to degree of implementation of the ERIP. A stratified sample of classrooms was then selected in which the highest implementers, considered the treatment group (type 1), were compared with the lowest implementers, representing the control group (type 2). Additional observations were recorded in each classroom of the sample to obtain a total of three observations per classroom. The mean of the three observation scores for a given classroom was the index of implementation for that classroom.

Standardized achievement tests in reading were used in September and May to ascertain what changes in pupil reading levels had occurred in the forty-three classrooms of the sample.

The process data (regarding classroom implementation) were compared, by analysis of variance and in terms of correlation coefficients, with the educational product data (regarding pupil achievement) as a means of determining the effectiveness of the ERIP.

### Development of the Observation Checklist

#### Rationale for the Use of a Sign System

After consideration of the various types of observation instruments described in the literature, a sign system

was selected as the most useful for this study. The sign system provided an instrument that was simple to use, required a minimum of observer training, and allowed observation and recording to be completed expeditiously. However, one of the strongest arguments in favor of selecting a sign type was cited by both Soar (1972) and Rosenshine (1973) that data from sign systems have proved to be more predictive of student achievement than data from other types of recording instruments.

#### Content of the Checklist

The stating of each item in the Observation Checklist of this study (Appendix B) was framed in the present tense to encourage the recording of a specific instance rather than an overall judgment of the observation period. Furthermore, the statement of each item attempted to keep the data on an objective, low inference level. The aim of the researcher was to state items with the specificity to maximize reliability of observation data and the generality to avoid becoming too narrowly restrictive in the behaviors to be observed. A number of the checklist items were accompanied by examples in order to help define specific kinds of observable behaviors that would be indicators for the item, yet would not prescribe limitations too narrow in scope for the item. The format adopted for the Observation Checklist in this study was similar to the structure of the TPOR (Brown, 1968).

The content of the checklist items was determined by the design of the ERIP and important aspects of its underlying

philosophy based on an understanding of reading pedagogy, teacher effectiveness, and learning theory. Part I of the Observation Checklist directed the observer to record the variety of approaches to reading instruction observed. The importance of variety of approach derived from the research conclusions that no one method of teaching reading was best for all learners, that teacher flexibility increased teacher effectiveness, and that a range of teaching-learning techniques accommodated differences among learners and learning tasks better than any one method could do. As indicators of variety of approach, the checklist employed specific behaviors characteristic of each of the three major approaches to reading instruction endorsed by the ERIP: language experience, individualized reading, and use of basal reader texts.

Part II of the Observation Checklist dealt with specific aspects of teaching language arts and reading skills which were found to be associated with greater cognitive achievement by pupils: the sequence in which specific skills are acquired, diagnosis of learning difficulties, direct teaching and reinforcing of specific skills, and the learning focus provided by the teacher. Items involving concept formation, oral language development, and practice in spelling and writing which have been found to affect reading progress and intellectual growth were also included.

In contrast to the first two parts of the checklist, which were concerned with program-specific, cognitive behaviors, Part III dealt with affective behaviors generally

applicable across programs and age levels. The interactions of teacher and pupil behaviors in the learning process were selected for observation on the dual basis that they were indicative of a favorable learning climate and that they were the affective factors which were thought to have the most consistent cognitive effects.

Since the items in the checklist were behaviors taught to teachers in the in-service phase of the ERIP and encouraged for regular use in classrooms, the items were considered a representative sample of the program. The total Observation Checklist served as an indicator of the degree to which the ERIP was being implemented in the classroom.

#### Validation Procedures

The validity of the checklist was tested by means of a number of criteria. First of all the three major parts of the checklist were selected according to areas indicated by theory and research to be important for reading achievement: variety of approach, direct teaching of skills, and warm classroom interaction.

To judge content validity, the checklist items were reviewed by personnel intimately involved with the ERIP: the designer-director of the ERIP and instructional consultants who helped develop and pilot the program. This panel of experts reviewed the checklist to judge and verify that it accurately translated objectives of the ERIP into representative observation items. Several modifications resulted from

their reviews and conferences with the researcher in studying and criticizing an early draft of the checklist (Appendix C).

For further refinement, the researcher engaged with a university professor of reading and language arts methods in a detailed, word-by-word analysis and overall assessment of the checklist. These discussions resulted in rewording and deletions from the somewhat unwieldy early draft of the instrument. The deletions imparted greater practicality to the observation instrument. The rewording had a net effect of replacing specific terms with generic terms, thus giving the checklist the possibility of wider application beyond the ERIP.

As a further validation study, the checklist was subjected to the criticism of advanced graduate students participating in a university seminar in research literature and special research problems in reading (Appendix D). Most members of this student group were in-service, experienced teachers, many of whom had advanced from regular classroom teaching into various roles of specialization in the teaching of reading. The graduate students studied the checklist as a class problem, submitted written responses, and discussed their suggestions, item by item, with the researcher. The two questions these critics applied to each checklist item were: A. If you were observing a class, would this item clearly specify to you what you are to look for? and B. If this item (or activity) were present, do you think you

would be able to recognize it? The seminar members suggested improvements for stating readily observable behaviors, and their suggested revisions clearly reflected their knowledge of reading instruction. The checklist items were again refined in wording and some were deleted--not because they were unimportant for a reading program but because they might have proved difficult for an observer to identify without prior knowledge of pupil assignments or individual pupil abilities, for instance. As a result of the foregoing validation reviews, items were restated with greater objectivity, and ambiguities were resolved. During the various steps of validation and refinement, the checklist was reduced from about sixty items to thirty-seven (Appendix B).

#### Reliability Studies

Reliability in using the checklist was established in field testing by the three instructional consultants. Field testing the use of the checklist in actual classroom observation served a dual purpose. First, the feasibility of the checklist items was tested by observation in natural classroom settings. Secondly, interobserver consistency was developed by the three consultants using the checklist in the same classroom simultaneously. The consultants independently recorded their observations on the checklist forms. At the completion of each observation, the observation records for that classroom situation were compared and analyzed. Where wide differences occurred among the observation records, such

items were modified in the way the items were stated, or an agreement was reached on what would be included as behavior for that item. Such analysis and discussion provided observer training which had the purpose of lowering the level of inference required of observers and also of increasing interobserver reliability. The checklist scores recorded by the three observers during the field testing are shown in Table 1. The graphic presentation in Figure 1 shows that all observers found the same classrooms relatively high or low in implementation.

#### Modification of the Recording Form

For convenience in classroom use, a Classroom Tally Sheet was devised (Appendix B). In this severely abbreviated form of the observation checklist, all of the items were included but were identified by a number and only partial spelling of each statement. Since this abridged form could not be easily read by anyone unfamiliar with the items, a person catching a glimpse of the Classroom Tally Sheet would not thereby be influenced to conform to the checklist behaviors. In addition, the extreme abbreviations made it necessary for the observers to know the items well in order to avoid confusion of the items during observations.

Thus the Observation Checklist was constructed to be consistent with the purposes of the ERIP and with theoretical and empirical bases of language arts instruction. Its reliability in classroom practice was tested. The feasibility of

Table 1

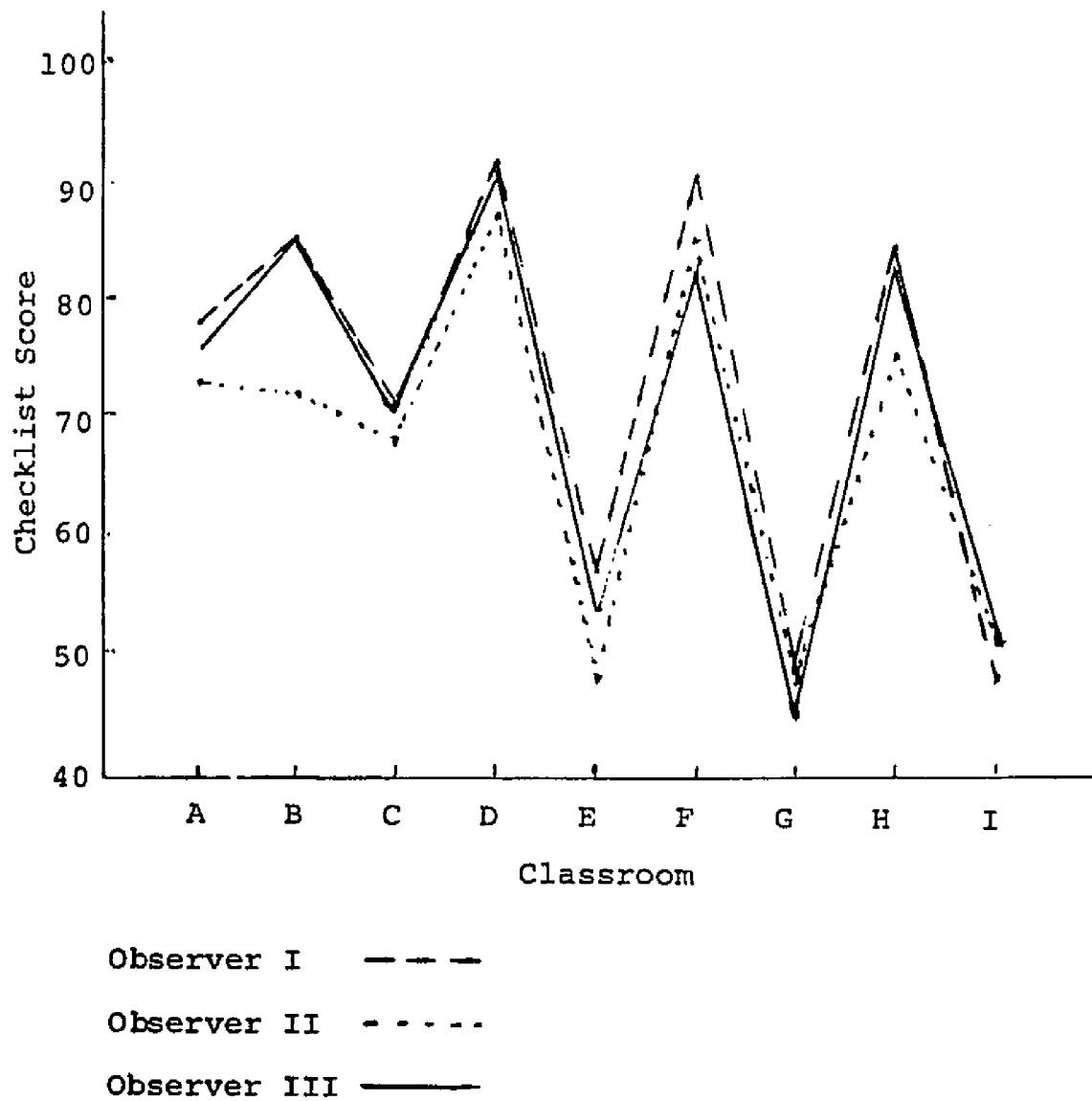
Reliability Studies: Total Observation Checklist Scores for  
Three Observers Independently Scoring the  
Same Classroom Simultaneously

Class- room	---Observer---		
	I	II	III
A	78	73	76
B	85	72	85
C	71	68	71
D	91	87	91
E	57	48	54
F	90	85	82
G	48	48	45
H	84	75	84
I	48	51	51

The correlation coefficients between checklist scores of simultaneous observer pairs were: I and II, .964; II and III, .951; I and III, .984. The overall correlation among the three observers in the reliability studies was thus .970. This average correlation was computed using Fisher's z function.



Figure 1  
Comparison of Simultaneous Observation Checklist  
Records of the Three Observers



the checklist was increased by utilizing the suggestions of experienced reading specialists, and its efficiency in use was augmented by the formulation of the Classroom Tally Sheet.

### Selection of the Stratified Sample

Fifteen elementary schools of the East Baton Rouge Parish school system were included in the ERIP (Appendix H). Of the total population of 249 teachers involved in the program, 118 were in classrooms which met the following criteria:

1. Only third, fourth, fifth, and sixth grade classrooms were used since standardized tests for these levels could be readily compared.

2. Only classes composed of students of the same grade level were included.

3. Any classroom whose teacher was a first-year teacher was not included in the study.

4. Any classroom whose teacher was having an acute personal or family problem known to the observers was not included in the study.

Each of the 118 classrooms that met the criteria for this study was ranked according to its observation checklist score based on an initial observation. From this ranking, a stratified sample of forty-eight classrooms was selected (Figure 2). The eight highest ranking (type 1) and the eight lowest ranking (type 2) in each of the three clusters were designated as the sample for this study (Table 2). Thus;

Table 2  
Checklist Scores for Initial Observation of Highest and Lowest  
Implementing Classrooms in Each Five-School Cluster

-----Highest - Type 1-----				-----Lowest - Type 2-----			
Teacher	School	Grade	Checklist Score	Teacher	School	Grade	Checklist Score
-----Cluster 1-----							
1	2	6	90	9	1	5	12
2	4	5	85	10	3	4	15
3	4	4	78	11	5	6	15
4	4	3	76	12	2	4	18
5	2	4	75	13	1	6	18
6	2	6	75	14	5	4	18
7	4	6	72	15	5	6	21
8	1	5/6	72	16	2	3	21
-----Cluster 2-----							
17	8	5	91	25	6	6	15
18	8	3	87	26	6	5	15
19	10	6	87	27	9	6	21
20	10	3	81	28	6	6	27
21	8	4	81	29	10	4	33
22	6	4	81	30	7	4	33
23	9	5	81	31	10	3	39
24	8	6	71	32	7	3	42
-----Cluster 3-----							
33	15	4	82	41	11	3	7
34	12	5	77	42	11	4	9
35	12	6	73	43	12	5	14
36	15	6	71	44	14	5	24
37	14	6	61	45	12	4	27
38	11	5	60	46	12	5	24
39	13	6	59	47	15	6	33
40	13	3	59	48	11	4	33

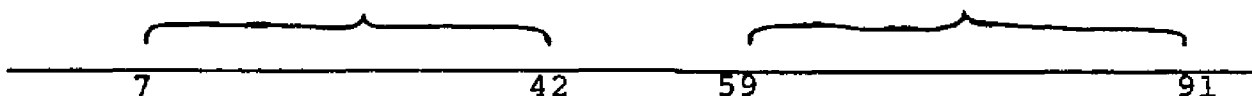
classrooms were selected, not by imposing the treatment on some and not on others, but by screening for those that had adopted most and least from the in-service education component of the program.

Figure 2

Selection of Sample Based on Checklist Scores  
for Initial Observation in 118 Classrooms  
Meeting the Criteria

Control Group - Type 2  
Eight lowest ranking in  
each of three clusters

Treatment Group - Type 1  
Eight highest ranking in  
each of three clusters



Total Checklist Score Ranges for Initial Observation

The twenty-four high implementing classrooms were compared with the twenty-four low implementing classrooms. Due to the method of sample selection, equal numbers of classrooms by grades or schools were not necessarily assured. Table 4 shows the distribution of classrooms in the sample by school, implementation type, and grade level. As Table 4 shows, the number of classrooms per school varied and schools also differed in having only type 1 teachers, only type 2 teachers, or a combination of both teacher types. In terms of grade level and implementation type, the sample presented the following distribution of classes:

Table 3  
Distribution of Classes

Grade	Type 1	Type 2	Total by Grade
3	3	3	6
4	5	7	12
5	5	4	9
6	<u>9</u>	<u>7</u>	<u>16</u>
Total by Type	22	21	43 (Total Sample)

The five classrooms lost from the sample were indicated by parentheses in Table 4. There were two high implementing and three low implementing teachers lost: two were third grades, one a fourth, one a fifth, and one a combination 5/6 grade. The combination classroom did not meet the criteria and was included in the sample by error. A third, a fourth, and a fifth grade were lost because either the pretests were not administered at the specified time or their scores were not available.

One third grade was deleted on the basis of race. In the data regarding race, the sample had only one teacher and ten students of race other than black and white, a nearly biracial population. To study the data on a biracial basis, one teacher, her eighteen students, and nine additional students of 'other race' were omitted since they could not properly be included in either black or white. The data of

Table 4  
Distribution of Classrooms in Sample  
by School, Type, and Grade

School	Type 1 - High Implementers Classroom	Grade	Type 2 - Low Implementers Classroom	Grade
1	(8	5/6) <sup>a</sup>	9 13	5 6
2	1 5 6	6 4 4	12 16	4 3
3			10	4
4	3 4 7 2	4 3 6 5		
5			11 14 15	6 4 6
6	22	4	25 26 28	6 5 6
7			(30 (32	4) <sup>b</sup> 3) <sup>b</sup>
8	17 18 21 24	5 3 4 6		
9	23	5	27	6
10	19 (20	6 3) <sup>c</sup>	29 31	4 3
11	38	5	41 42 48	3 4 4
12	34  35	5  6	45 (43 46	4 5) <sup>b</sup> 5
13	39 40	6 3		
14	37	6	44	5
15	33 36	4 6	47	6

Notes: a. Combination grade level does not meet criteria.  
b. Unacceptable pretest.  
c. Other race.

the study included an actual sample of 43 teachers and the 994 students who were members of their classes from pretesting through posttesting.

### Collection of Observation Data

The Observation Checklist was studied by the three observers to thoroughly familiarize themselves with the items to be observed. In actual observation, the observers used the Classroom Tally Sheet (Appendix B). The field testing and reliability studies had provided the observers with guidance and practice in using the observation instrument reliably. The high and low implementers were determined on the basis of an initial observation in each eligible classroom of the ERIP. During the school year, an instructional consultant completed two additional observations in each of the classrooms designated for this study.

A single observation period entailed observing the classroom activities to determine the presence of each item on the list in three discrete time intervals. In the first time interval, each item was recorded in column one as a "1" if the behavior was found, or no tally was marked if that behavior was not present. A time interval required about seven to ten minutes to make a determination for each checklist item. Immediately upon completion of one time interval of observation, a second time interval was carried out and recorded in column two in the same way, followed by a third

time interval using column three. This sequence of three time intervals of observing and recording constituted one classroom observation of approximately one half hour. The number of time segments in which an item occurred resulted in a quantitative measure of its frequency of occurrence. For each observation, an item on the checklist could accrue a score of 0, 1, 2, or 3. The score for an observation was the grand total of the thirty-seven item totals. The range of the grand total could be as low as 0 or as high as 111.

One of the problems encountered in direct observation has been the change in classroom atmosphere caused by the presence of the stranger who is observing. Since the observers in this study were the instructional consultants who regularly worked in these classrooms, no unaccustomed stranger was introduced into the learning environment. Thus the use of a usual participant-observer avoided a novel intrusion which has been a frequent disadvantage of direct classroom observation.

The use of the instructional consultants as observers also mitigated another problem mentioned by Withall (1949). He noted that observers required both much knowledge of subject matter and of behavior observation making observer training a highly technical problem. Knowledge of subject matter was a requirement for appointment to the consultant position. Observation of classroom behavior was part of the daily task of the consultant. Due to familiarity with



language arts instruction and behavior observation in the conduct of their in-service function, the consultants required only the additional knowledge of the items and scoring system specific to this Observation Checklist. Thus the observer training was considerably simplified.

The Classroom Tally Sheets recording the three observations in each selected classroom were collected by the researcher. All totals and grand totals were rechecked for accuracy and the mean of the three grand totals was calculated to find the index of implementation for each classroom. Finally, the observation data was tabulated.

#### Administration of Achievement Tests

Science Research Associates (SRA) Achievement Test Series in reading was used as the measure of reading ability. In grades 3 and 4, Reading--Forms E and F/Primary II were used for pretest and posttest, respectively. In grades 5 and 6, Multilevel Reading Test, Forms C and D were used for pretest and posttest, respectively. The pretesting was conducted between September 8 and 12, 1975, and the posttesting was carried out during the period of May 4 to 14, 1976. The SRA tests were administered by the classroom teachers to their own students so that the test setting would be as normal as possible for the students. In addition to the manual from the publisher included in each teacher's test packet, written instructions regarding administration were provided for each

teacher in the ERIP (Appendix F). The amount of change between pretest and posttest scores adjusted for initial ability was calculated by computer for each pupil. The mean change of each class as a unit was used to determine amount of change in reading ability for the year. The class mean change in reading achievement was compared with the index of implementation of the ERIP for each classroom.

#### Treatment of the Data

Checklist data, reading achievement test results, and student and teacher information were compiled by the researcher and recorded on coding sheets. The data were then transferred to computer cards for processing. Using the Statistical Analysis System (SAS), analyses of variance procedures were applied to the data (Appendix G) to determine which variables had a significant effect on raw posttest scores and growth in reading, and correlation coefficients were obtained to determine significant relationships among the variables.

#### Summary

This study involved the development, validation, and use of a checklist for classroom observation. After conducting reliability studies and observer training sessions, three observers applied the checklist in each classroom which met the criteria of this study. On the basis of this initial

observation, stratified samples of twenty-two high implementing classrooms and twenty-one low implementing classrooms of the ERIP were selected to serve as treatment and control groups, respectively. Additional observations were carried out in the selected classrooms to provide a larger sample of process data. Pre- and posttest scores of the SRA achievement tests in reading, administered in September, 1975, and May, 1976, were used as the measure of educational outcome. Analyses of variance were used to determine the effects of the variables on posttest scores and gains in reading achievement. The comparisons which were made between the changes in reading achievement of pupils in the high implementing classrooms and similar changes of pupils in low implementing classrooms provided a measure by which to evaluate the reading program being observed.

## Chapter 4

### PRESENTATION AND ANALYSIS OF DATA

The problem addressed in this study was: How was pupil reading achievement related to the degree of teacher implementation of the Elementary Reading Improvement Program (ERIP)? There were four questions subsumed under the stated problem:

1. How was pupil reading achievement related to the classroom practices utilized in variety of approaches to reading instruction?

2. How was pupil reading achievement related to the classroom practices utilized in diagnosis and development of specific reading skills?

3. How was pupil reading achievement related to classroom interaction?

4. How was pupil reading achievement, with respect to pupil sex, pupil race, teacher race, and grade level, related to the index of implementation?

The population for this study was a stratified sample of twenty-two classes found to be high implementers of the ERIP and twenty-one low implementing classes involving a total of forty-three teachers and their 994 students.

Analyses of variance were used to find answers to the stated questions. Comparisons were made between the index of implementation for each class with the class mean of the difference

between pretest and posttest reading achievement scores adjusted for initial variations in pretest scores. The index of implementation was defined for this report as the mean of the total checklist scores for a given classroom. This chapter discusses the data in relation to each of the questions.

### Reading Measures and Implementation Measures

The data from this study showed highly significant positive correlations ( $p < .01$ ) in two areas: between pretest and posttest reading scores, and among Observation Checklist scores on Parts I, II, and III, and total checklist scores in any combination of those four scores (Table 5). In regard to reading achievement, the students who had lower pretest scores also had lower posttest scores, and conversely, higher pretest scorers had higher posttest scores. However, comparing the amount of gain in reading achievement with pretest scores (Table 6) showed highly significant negative correlations ( $p < .01$ ) for the forty-three classrooms. The correlation of  $-.42$  (Table 5) indicated that lower pretest scores were associated with greater gains than were the higher pretest scores. Reading scores as well as implementation scores were reported using the class as a unit.

The Observation Checklist scores indicated that implementation of any part of the checklist correlated to a highly significant degree ( $p < .01$ ) with the implementation of each other part of the checklist and with the total checklist

Table 5

Correlation Coefficients Showing Relationships  
Among Checklist Scores and Among Reading  
Achievement Measures for Total Sample

	Observation Checklist Scores			Reading Test Scores		
	Part II	Part III	Total	Pre Raw	Post Raw	Raw Gain
Part I	0.89**	0.93**	0.98**	0.01	-0.02	-0.09
Part II		0.90**	0.94**	-0.09	-0.11	-0.03
Part III			0.98**	-0.04	-0.07	-0.07
Total				-0.03	-0.06	-0.07
Pre Raw					0.95**	-0.42** <sup>a</sup>
Post Raw						-0.13

\*\* =  $p < .01$

<sup>a</sup>Henry E. Garrett, Statistics in Psychology and Education, (6th ed.,; McKay, 1966), p. 201. As Garrett shows, even a correlation of absolute value less than .40 is significant at the .01 level for sample size in excess of forty.

Table 6

Frequency Distribution of Classrooms According to  
Mean Pretest and Mean Gain for Total Sample

Raw Score Mean Gain	Raw Score Mean Pretest					
	0-9	10-19	20-29	30-39	40-49	50-59
15-17	1					
12-14		2	1			
9-11		1	4			
6-8			6	9	1	
3-5		1	1	5	3	1
0-2			2		1	
-3--1			1	1		

score as shown in Table 5. Since there was such a close match among these scores, results were discussed in terms of the index of implementation. Thus the results using total checklist implementation scores were similar to findings for the first three questions which related to Parts I, II, and III, respectively, of the checklist. Checklist scores for each classroom in the study were presented in Appendix G.

#### Effects of Variables on Reading Achievement

In regard to the general problem of this study: How was pupil reading achievement related to the degree of teacher implementation of the ERIP?, the data showed an overall negative relationship between reading achievement gains and levels of implementation of the ERIP (Table 7).

Table 7

SRA Reading Achievement in Raw Score Measures  
for High and Low Implementing Classrooms

Implemen- tation Type	N	Mean Pre- Test	Mean Post- Test	Mean Gain
High	516	32.3	37.8	5.5
Low	478	32.4	39.3*	6.9*

\* = significantly higher at  $p < .05$

The questions posed in this study and the data related to each question follow:



1. How was pupil reading achievement related to the classroom practices utilized in variety of approaches to reading instruction? The data of this study indicated that overall reading achievement measures of the sample population were negatively related, to a significant degree ( $p < .05$ ), to variety of approaches used in classroom instruction as measured by Part I of the Observation Checklist (Tables 5 and 7).

2. How was pupil reading achievement related to the classroom practices utilized in diagnosis and development of specific reading skills? In answer to question 2, gains in reading achievement scores in this study were negatively related, to a significant degree ( $p < .05$ ), to diagnosis and instruction in specific reading skills as determined by Part II of the Observation Checklist (Tables 5 and 7).

3. How was pupil reading achievement related to classroom interaction? Similarly for question 3, measures of reading achievement gains in this evaluation were shown to be negatively related, to a significant degree ( $p < .05$ ), to the measures of classroom interaction used in Part III of the Observation Checklist (Tables 5 and 7).

Comparing mean reading gains of two distinct categories--classrooms implementing the ERIP to a high degree and low implementing classrooms--reading gains were significantly greater ( $p < .05$ ) in low implementing classrooms (6.9) than in high implementing classrooms (5.5) (Table 7). Likewise, the mean of posttest raw scores was significantly

greater ( $p < .05$ ) for the low implementing group (39.3) than for the high implementers (37.8).

Analyses of the effects of race, sex, and grade on reading achievement were used to deal with question 4, which asked:

4. How was reading achievement, with respect to pupil sex, pupil race, teacher race, and grade level, related to the index of implementation? In the total sample of this study, white students made gains greater to a highly significant degree ( $p < .01$ ) than black students (Table 8), and girls made greater gains than boys (Table 9) although these differences were not significant. Table 10, representing the interaction of these two factors, showed the following relationships: white girls made the greatest gains (7.83); white boys ranked slightly lower (7.76); below a wider gap, black girls ranked next (5.51); and black boys made smaller gains (3.70) than any of the other three groups.

Student gains in reading for each grade level were shown in Table 11 by number of classrooms at grade level and then classified according to high and low implementation groups for further comparison at each grade level. According to the data as shown in Table 11, grade level makes a difference to a highly significant degree ( $p < .01$ ). However, the differences due to grade level were not the same among the high and low implementers. The findings shown in Table 11 seemed to exhibit no logical pattern. It was also noted that

Table 8

SRA Reading Achievement in Raw Score Measures  
Classified by Student Race

Student Race	N	Mean Pre- test	Mean Post- test	Mean Gain
White	560	32.2	40.1**	7.8**
Black	434	32.4	37.0	4.6

\*\* = significant at  $p < .01$ .

Table 9

SRA Reading Achievement in Raw Score Measures  
Classified by Student Sex

Student Sex	N	Mean Pre- test	Mean Post- test	Mean Gain
Boys	467	32.4	38.1	5.7
Girls	527	32.3	39.0	6.7

Table 10

SRA Reading Achievement in Raw Score Measures  
Classified by Student Sex by Student Race

Student Sex	N	-----White Students-----			N	-----Black Students-----		
		Mean Pre- test	Mean Post- test	Mean Gain		Mean Pre- test	Mean Post- test	Mean Gain
Boys	267	32.4	40.1	7.76	200	32.4	36.1	3.70
Girls	293	32.4	40.2	7.83	234	32.4	37.9	5.51

Table 11

SRA Reading Achievement in Raw Score Measures  
Classified by Grade Level and Implementation Type

Grade	---Total Sample---				-High Implementers				-Low Implementers--			
	N	Mean Pre- test	Mean Post- test	Mean Gain	N	Mean Pre- test	Mean Post- test	Mean Gain	N	Mean Pre- test	Mean Post- test	Mean Gain
3	6	32.4	40.7**	8.3**	3	32.4	39.6**	7.2**	3	32.4	41.8**	9.4**
4	12	32.3	37.2**	4.9**	5	32.4	37.6**	5.2**	7	32.3	36.8**	4.5**
5	9	32.4	38.2**	5.8**	5	32.4	36.0**	3.6**	4	32.3	40.4**	8.1**
6	16	32.3	38.1**	5.8**	9	32.4	38.3**	5.9**	7	32.4	38.0**	5.6**

\*\* =  $p < .01$

N = Classrooms

the numbers of cases (class means) represented by each cell were relatively small for statistical purposes (Tables 3 and 11). Students in the high implementing classrooms showed slightly larger reading achievement gains than the low implementers at grades 4 and 6. In contrast, students in the low implementing classrooms showed considerably larger reading achievement gains than the high implementers at grades 3 and 5 (Figure 3).

The influence of teacher race was shown in Table 12. The differences due to teacher race were not statistically significant. The highest mean gains in reading achievement test scores were reported for students in low implementing classrooms with white teachers (7.2). Next in rank were the reading gains attributed to students in low implementing classrooms with black teachers (6.7); then, students in high implementing classrooms with black teachers (6.2); least gains were indicated by the reading scores reported for high implementing classes with white teachers (4.7).

The results comparing interaction of teacher race and student race (Table 13) showed no significant differences. The data indicated that the greatest mean gain (7.9) in reading achievement scores were made by white students with white teachers. White students with black teachers ranked next highest scoring a mean gain of 7.7. Black students with black teachers showed a mean gain of 5.2. Black students with white teachers showed the least mean gain (4.0). The

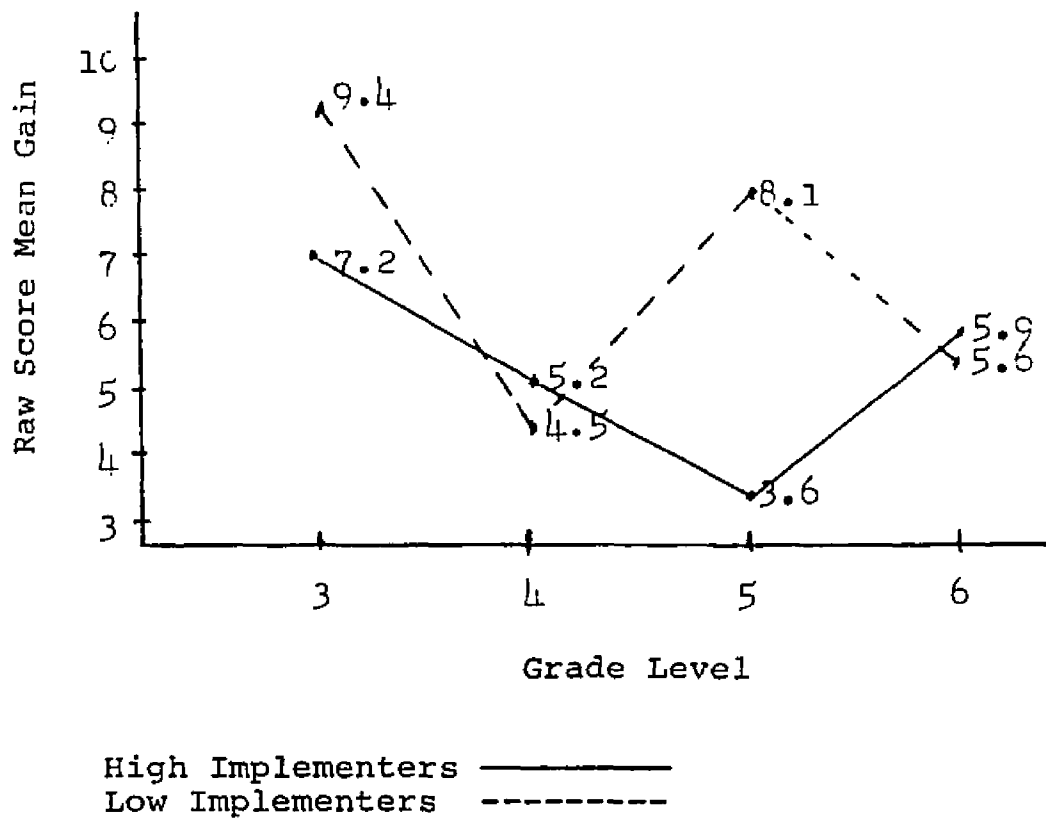


Figure 3  
Comparison of Mean Gains of High and  
Low Implementers by Grade Level

Table 12

SRA Reading Achievement in Raw Score Measures Classified  
by Teacher Race and Implementation Type

Teacher Race	N	-----Total Sample-----			N	----High Implementers----			N	-----Low Implementers-----		
		Mean Pre- test	Mean Post- test	Mean Gain		Mean Pre- test	Mean Post- test	Mean Gain		Mean Pre- test	Mean Post- test	Mean Gain
White	631	32.4	38.3	5.9	377	32.4	37.1	4.7	254	32.3	39.5	7.2
Black	363	32.3	38.8	6.5	139	32.4	38.6	6.2	224	32.3	39.0	6.7

Table 13

SRA Reading Achievement in Raw Score Measures  
Classified by Student and Teacher Race

Stu- dent Race	-----White Teachers-----				-----Black Teachers-----				-----All Teachers-----			
	N	Mean Pre- test	Mean Post- test	Mean Gain	N	Mean Pre- test	Mean Post- test	Mean Gain	N	Mean Pre- test	Mean Post- test	Mean Gain
White	403	32.3	40.2	7.9	157	32.3	40.0	7.7	560	32.3	40.1**	7.8**
Black	228	32.4	36.4	4.0	206	32.4	37.6	5.2	434	32.4	37.0**	4.6**
All	631	32.4	38.3	5.9	363	32.3	38.8	6.5	994 <sup>a</sup>	32.4 <sup>a</sup>	38.3 <sup>a</sup>	5.9 <sup>a</sup>

\*\* =  $p < .01$ .

N = Number of students (16 black teachers; 27 white teachers).

<sup>a</sup>Total sample of this study.



trend appears to be for black students to do better relative to white students when they have black teachers rather than white teachers. However, since the figures are not statistically significant, they can only suggest a possible trend.

Mean reading gains for individual classes were compared with unadjusted mean pretest scores by implementation type in the frequency distributions in Appendix I. As was the case for the total sample, the lower achieving classes at pretest tended to be associated with greater reading gains in classes of low implementation. This phenomenon was not as clearly evident in the relationship between pretest and gains for the high implementing classes. These findings for the low achievers seemed to parallel Dykstra's (1967) reports that less mature students in the First Grade Reading Studies profited more from a basal reader program in contrast to his higher achievers who did better in the less narrowly structured language experience approach. It may also be true that the low achievers were still at the stage of gaining mostly simple-concrete skills which Soar (1970, 1972, 1976) showed were better learned in tightly controlled situations such as low implementing classes tended to be.

Examination of the data shown in Appendix G revealed that some individual classes reported unexpectedly high achievement gains which were difficult to explain. These expansive gains exerted a strong effect on the mean gain in reading achievement for their implementation type and grade level.

### Summary

In the classrooms with low implementation of ERIP behaviors, reading achievement was significantly greater ( $p < .05$ ), in terms of both raw mean posttest scores and raw mean gains, than in the high implementing classrooms. Although grade level had a highly significant effect ( $p < .01$ ) on reading achievement, the patterns for high and low implementers from grade to grade were pronounced and diverse. The means of categories containing only a few classes may have been overly affected by the extreme results reported for a few individual classes. Student race had a highly significant effect ( $p < .01$ ) on reading achievement. The data of this study did not show that the other variables tested exerted significant effects.

## Chapter 5

### SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

This evaluation of the Elementary Reading Improvement Program of the Baton Rouge (Louisiana) public schools, during 1975-76, was conducted to find out if concentration of the program in a classroom would improve reading achievement for the students in that classroom. The expectation that high implementation would increase student achievement in reading was not supported by the data in this study. Since the ERIP was designed on sound bases of theory and research regarding reading and language arts instruction, it did not seem reasonable to question the soundness of the pedagogical principles or program design involved. Other possible explanations were sought in relation to the assessment of implementation of the program, or the reading achievement measure of the students, or the in-service help given to the teachers.

#### Summary

The problem to be investigated in this research was: How was pupil reading achievement related to the degree of teacher implementation of the Elementary Reading Improvement Program? The questions to be answered were:

1. How was pupil reading achievement related to the classroom practices utilized in variety of approaches to reading instruction?

2. How was pupil reading achievement related to the classroom practices utilized in diagnosis and development of specific reading skills?

3. How was pupil reading achievement related to classroom interaction?

4. How was pupil reading achievement, with respect to pupil sex, pupil race, teacher race, and grade level, related to the index of implementation?

This study involved the development, validation, and use of a checklist for classroom observation. After conducting reliability studies and observer training sessions, three observers applied the Observer Checklist in each classroom which met the criteria of the study. On the basis of this initial observation, a stratified sample of twenty-two high implementing classrooms and twenty-one low implementing classrooms of the ERIP were selected to serve as treatment and control groups, respectively. The population for this study included forty-three teachers and the 994 students who were members of their classes from pretesting in September through posttesting in May. Additional observations were carried out in the selected classrooms to provide a larger sample of process data. Pre- and posttest scores of the SRA achievement tests in reading were used as the measure of educational

outcome. Analyses of variance procedures were applied to the data to determine which variables had a significant effect on raw posttest scores and growth in reading. Correlation coefficients were used to determine the levels of significance of relationships among the variables. The comparisons which were made between the changes in reading achievement of pupils in the high implementing classrooms and similar changes of pupils in low implementing classrooms provided an evaluation of the reading program being observed.

### Findings

The findings of this investigation showed differences highly significant at the .01 level of confidence unless otherwise noted. Regarding the main problem of this study, mean reading gains were significantly greater ( $p < .05$ ) for low implementing classrooms (6.9) than for high implementing classrooms (5.5). The following data applied to the main problem:

- a. The Observation Checklist scores of Parts I, II, and III, and total checklist scores, in any combination, correlated to a highly significant degree.
- b. Pretest and posttest reading scores, using the mean of the class as a unit, showed highly significant correlations.
- c. The amount of mean gain in pupil reading achievement was negatively correlated, to a highly significant degree, with mean pretest reading scores.

1. In the findings for question 1, pupil reading achievement showed a negative relationship to a significant degree ( $p < .05$ ) to classroom use of variety of approaches to reading instruction according to the data in this study.

2. In the findings for question 2, pupil reading achievement showed a negative relationship to a significant degree ( $p < .05$ ) to classroom use of diagnosis and development of specific reading skills according to the data in this study.

3. In the findings for question 3, pupil reading achievement showed a negative relationship to a significant degree ( $p < .05$ ) to classroom interaction according to the data in this study.

4. The findings for question 4 indicated that pupil reading achievement was affected in differing ways by the variables tested in this study as follows:

a. Pupil race showed highly significant differences in pupil reading achievement gains.

b. Pupil sex made no significant differences in pupil reading achievement gains.

c. In the interactions between pupil race and pupil sex, there were no significant differences in pupil reading achievement gains.

d. The interaction between grade level and implementation type produced strong and divergent results. In grades 4 and 6, high implementing classrooms showed slightly larger gains than low implementers at a highly significant

level of confidence. In grades 3 and 5, low implementing classrooms showed considerably greater gain in reading achievement scores than the high implementers at a highly significant level of confidence.

e. Teacher race made no significant differences in pupil reading achievement gains.

f. The interaction between teacher race and implementation type showed no significant differences in pupil reading achievement gains.

g. The interactions between teacher race and pupil race showed no significant differences in pupil reading achievement gains.

### Conclusions

In the light of the data presented and the limitations imposed by the study, the following conclusions appear to be indicated in regard to the main problem:

Since pupil reading achievement gains were significantly greater in the low implementing classrooms, the low implementing classrooms appeared to be of more advantage than the high implementing classrooms for improving pupil reading achievement according to the standardized tests used.

a. Since all three parts of the Observation Checklist scores correlated closely with each other and with the total checklist score, results could be discussed in terms of the total checklist score average called the index of implementation.

b. Since pretest and posttest mean scores had a highly significant correlation, pretest scores could be considered good predictors of posttest scores.

c. Students who were lower achievers initially made greater progress to a highly significant degree than those with higher pretest scores.

These conclusions were suggested by the questions to be answered:

1. Greater use of variety of approaches to reading instruction was associated with lower posttest scores and less gain in achievement according to the standardized reading test used.

2. More classroom implementation of diagnosis and direct teaching of specific reading skills was related to smaller achievement gains according to the standardized reading tests used.

3. In classrooms where more pupil-teacher interaction was observed, reading achievement gains were less as measured by the standardized reading tests.

4. The variables studied had diverse effects on pupil reading achievement as follows:

a. Apparently, no significant differences in pupil reading achievement gains were exerted by the variables of pupil sex nor teacher race, nor by the interactions between pupil race and pupil sex, between teacher race and implementation type, nor between teacher race and pupil race.



b. Grade level and the interaction between grade level and implementation type appeared to make highly significant differences in pupil reading achievement gains. The patterns of differences were not the same for the two implementation levels. The high implementation classes seemed to offer a small advantage to fourth and sixth graders and a great disadvantage to third and fifth graders. Study of the data presented no discernible explanation for the sharp contrasts from grade to grade.

#### Recommendations

1. Implementational and observational evaluation of instructional programs should be pursued, as Rosenshine (1973) and Soar (1970) have suggested, but with renewed efforts to improve and refine the measurement of both process and product.

2. Improvement of observation measures should attempt greater quantification such as amount or percentage of time per day spent in various kinds of instructional activities. Research involving time measurement has been reported by Harris (1966, 1968), McDonald (1976a,b,c,d), Elias (1976a,b,c), Coles (1976), Conner and Eisenberg (1966), and Durkin (1978).

3. Further observational studies should attempt to determine the match between individual pupil needs and the specific approaches each pupil experiences. Various approaches to this problem have been discussed by Rosenshine (1973) and Coles (1976) among others. As a companion to the investigation

of what pupils experience, it would be useful to determine to what extent and on what basis teachers used specific approaches or techniques for particular types of learners.

4. Careful study should be made in process-product studies to select a product measure and analysis which can provide a very close match to the objectives and skills included in the classroom instructional process. This suggestion might be implemented by using only the scores of criterion-referenced test items for each student which match objectives and skills listed by the teacher on a questionnaire as having been taught during that year to that student. A number of studies describing the use of teacher reports or diaries regarding planning and instructional activities might provide help in designing such a teacher questionnaire: SMSG

(Torrance, 1966), BTES (McDonald, 1976a,b,c,d; Calfee, 1976a, b,c; Elias, 1976a,b,c), CRAFT (Harris, 1966, 1968).

5. If possible, product measures should also include low inference observational records of performance items which cannot be measured by paper-and-pencil tests.

6. Further studies of grade level comparisons, possibly similar to grade level studies of Powell (1968), McDonald (1976a,b,c,d), or Wallen (1966), should be conducted to determine if grade level differences regularly exert such strong and erratic effects on learners in individualized and non-individualized programs.

7. Further attempts should be made to identify variables that cause some children (Coles, 1976) and/or some classes to make unexpectedly high gains.

8. A means of assessing the long-range effects of the ERIP and similar programs would be a valuable addition to the type of on-site evaluation presented in this study.

9. More cultural and linguistic studies for pre-service and in-service elementary teachers as suggested by Shuy (1971) might increase teachers' understanding of other cultures and of nonstandard English and their effects on learning to read in the experience of their pupils of variant cultures.

10. As further correlational studies are able to help tease out promising elements of instruction, such factors should then be tried out in experimental or pilot studies and, beyond that, incorporated into teacher education and universal classroom use whenever warranted.

## BIBLIOGRAPHY

## SELECTED BIBLIOGRAPHY

### BOOKS

- Bennett, Neville, et al. Teaching Styles and Pupil Progress. Cambridge: Harvard University Press, 1976.
- Brown, Bob Burton. The Experimental Mind in Education. New York: Harper & Row, 1968.
- Chall, Jeanne S. Learning to Read: The Great Debate. New York: McGraw-Hill, 1967.
- Dewey, John. Democracy and Education. New York: MacMillan, (1st ed. 1944), 1961.
- Dunkin, M. J., and B. J. Biddle. The Study of Teaching. New York: Holt, Rinehart and Winston, 1974.
- Garrett, Henry E. Statistics in Psychology and Education. New York: David McKay, 1966.
- Lavatelli, Celia S. Piaget's Theory Applied to an Early Childhood Curriculum. Cambridge: Center for Media Development, 2d ed., 1973.
- Simon, Anita, and E. Gil Boyer, eds. Mirrors for Behavior: An Anthology of Classroom Observation Instruments. Philadelphia: Research for Better Schools, 1967.

### ARTICLES, REPORTS, AND PAPERS

- Ackerman, W. I. "Teacher Competence and Pupil Change," Harvard Educational Review, 24 (1954), 273-289.
- Amidon, E. J., and N. A. Flanders. "The Effects of Direct and Indirect Teacher Influence on Dependent-Prone Students Learning Geometry," Journal of Educational Psychology, 52 (April 1966), 54-59.
- Bissell, J. "Implementation of Planned Variation in Head Start, I. Review and Summary of the Stanford Research Institute Interim Report: First Year of Evaluation." Washington, D. C.: U. S. Dept. of Health, Education and Welfare, Office of Child Development, 1971.

- Bond, Guy L. "The Coordinated Phases of the Reading Study," Progress Report presented at the Annual Conference of the International Reading Association, Dallas, May, 1966.
- \_\_\_\_\_, and R. Dykstra. "The Cooperative Research Program in First Grade Reading Instruction," Reading Research Quarterly, 2 (1967), 1-142.
- Brown, Bob Burton. "Experimentalism in Teaching Practice," Journal of Research and Development in Education, 4 (1970), 14-22.
- Bruner, J. "The Course of Cognitive Growth," American Psychologist, 19 (1964), 1-15.
- Calfee, Robert C., and Kathryn Hoover Calfee. "Reading and Mathematics Observation System: Description and Measurement of Time Usage in the Classroom" (RAMOS), Beginning Teacher Evaluation Study: Phase II, 1973-74 (BTES), Vol. III, Chapter 2, 1976a, ED 127 367.
- \_\_\_\_\_. "Reading and Mathematics Observation System," Final Reports, Princeton: Educational Testing Service, 1975-76, 1976b, ED 131 115,
- \_\_\_\_\_. Special Report, Journal of Teacher Education, XXVII (Winter, 1976c) 4, 323-325.
- Calvin, A.D., et al. "The Effect of Intelligence and Social Atmosphere on Group Problem Solving Behaviour," Journal of Social Psychology, 45 (1957), 61-74.
- Coles, Gary J., et al. "Impact of Educational Innovation on Student Performance: Project Methods and Findings for Three Cohorts," Project LONGSTEP Final Report, Vol. I, AIR, 1976, (Longitudinal Study of Educational Practices) ED 132 177 (Also see USOE Executive Summary and ED 133 369).
- Conners, C. K., and L. Eisenberg. "The Effect of Teacher Behavior on Verbal Intelligence in Operation Head Start Children." (Head Start Contract No. 510, U. S. Office of Economic Opportunity.) Baltimore: School of Medicine, Johns Hopkins University, 1966.
- Durkin, D. "What Classroom Observations Reveal about Reading Comprehension Instruction." Technical Report No. 106. Champaign: Center for the Study of Reading, October, 1978, ED 162 259.
- Dykstra, Robert. "Continuation of the Coordinating Center for First Grade Reading Instruction Programs." Final Report, September 1967, ED 015 865.

- Dykstra, Robert. "Classroom Implications of the First Grade Reading Studies." April 1968, ED 022 626.
- Elias, Patricia, et al. "Reports of Teachers about Their Mathematics and Reading Instructional Activities," Beginning Teacher Evaluation Study: Phase II, 1973-74 (BTES), Vol. V, Chapter 5, 1976a, ED 127 374.
- \_\_\_\_\_. "Reports of Teachers" (BTES), Berkeley: Educational Testing Service, AERA Paper, 1976b, ED 131 116.
- \_\_\_\_\_. "Instructional Activities as Reported by Teachers," Special Report, Journal of Teacher Education, XXVII, No. 4 (Winter, 1976c), 326-328.
- Fattu, Nicholas A. "Effectiveness: An Elusive Quality," Education Digest, 27 (1962), 24-26.
- Flanders, Ned A. "Teacher Influence, Pupil Attitudes, and Achievement," Cooperative Research Monograph No. 12 OE-25040, Washington, D. C.: Government Printing Office, 1965.
- \_\_\_\_\_, and Anita Simon. Teacher Effectiveness, in R. L. Ebel (Ed.), Encyclopedia of Educational Research. 4th ed. Toronto: MacMillan, 1969, 1424-1434.
- Fortune, J. C. "The Generality of Presenting Behaviors in Teaching Preschool Children," Unpublished paper, 1966.
- \_\_\_\_\_. "A Study of the Generalities of Presenting Behaviors in Teaching." Project No. 6-8468, U. S. Office of Education. Memphis: Memphis State University, 1967.
- Fowler, B. D. "Relations of Teacher Personality Characteristics and Attitudes to Teacher-pupil Rapport and Emotional Climate in the Elementary Classroom," Dissertation, University of South Carolina, 1962.
- Furst, Norma. "The Multiple Languages of the Classroom." Paper read at AERA, New York, 1967.
- \_\_\_\_\_, and Russell A. Hill. Systematic Classroom Observation, in Lee Deighton (Ed.), Encyclopedia of Education. New York: MacMillan, 1971, 168-183.
- Gage, N. L. "Desirable Behaviors of Teachers," Urban Education, 1 (1965), 85-95.
- Grannis, J. Columbia Classroom Environments Project, Fifth Progress Report, Contract No. OEC-O-71-6593, USOE, May, 1972. ED 080 561 (Also see ED 123 292 Final Report).

- Grimes, J. W., and W. Allinsmith. "Compulsivity, Anxiety, and School Achievement," Merrill Palmer Quarterly, 7 (1961), 247-271.
- Howsam, R. B. Who's a Good Teacher? Problems and Progress in Teacher Evaluation. California Teachers Association, 1960.
- Harris, A. J., and B. L. Serwer. "Comparison of Reading Approaches in First Grade Teaching with Disadvantaged Children (CRAFT Project): Instructional Time in Reading Research," Reading Research Quarterly, II (Fall 1966), 27-56.
- \_\_\_\_\_, et al. "A Continuation of the CRAFT Project: Comparing Reading Approaches with Disadvantaged Urban Negro Children in Primary Grades," Cooperative Research Project No. 5-0570-2-12-1. New York: Division of Teacher Education of the City University of New York, 1968, ED 020 297.
- \_\_\_\_\_, and C. Morrison. "The CRAFT Project: A Final Report," Reading Teacher, 22 (January 1969), 335-340.
- Johns, J. P. "Relationship between Teacher Behaviors and Incidence of Thought-provoking Questions," Dissertation, University of Michigan, 1966.
- Lambert, Nadine M., et al. "Anecdotal Processing to Promote Learning Experience (APPLE): Observation Variables as Measures of Teacher Performance," Beginning Teacher Evaluation Study: Phase II, 1973-74 (BTES), Vol. III, Chapter 1, 1976a, ED 127 366.
- \_\_\_\_\_. "APPLE Observation Variables," (BTES), AERA Paper, 1976b, ED 131 114.
- \_\_\_\_\_. "APPLE Observation Variables," (BTES), Special Report, Journal of Teacher Education, XXVII, No. 4 (Winter, 1976c), 320-323.
- McDonald, Frederick J., et al. "Beginning Teacher Evaluation Study: Phase II, 1973-74" (BTES), Final Report, Vols. I-V, 5 and Summary, 1976a, ED 127 364-75.
- \_\_\_\_\_. "A Report on the Results of Phase II of the BTES: The Effects of Teaching Performances on Student Learning," AERA Paper, 1976b, ED 131 117.
- \_\_\_\_\_. "A Report on the Results of Phase II of the Beginning Teacher Evaluation Study: An Overview," Special Report, Journal of Teacher Education, XXVII, No. 4 (Winter, 1976c), 315-334.



- McDonald, Frederick J., and Patricia Elias. Beginning Teacher Evaluation Study Phase II, 1973-74," Executive Summary Report. Princeton: Educational Testing Service, 1976d.
- McNeil, John D., and James W. Popham. The Assessment of Teacher Competence, in R. M. W. Travers (Ed.), Second Handbook of Research on Teaching. (AERA) Chicago: Rand McNally, 1973, 221-226.
- Medley, Donald M., and Harold E. Mitzel. "A Technique for Measuring Classroom Behavior," Journal of Educational Psychology, 49 (1958), 86-92.
- \_\_\_\_\_. "Some Behavioral Correlates of Teacher Effectiveness," Journal of Educational Psychology, 50 (1959), 239-246.
- \_\_\_\_\_. Measuring Classroom Behavior by Systematic Observation, in N. L. Gage (Ed.), Handbook of Research on Teaching. Chicago: Rand McNally, 1963, 247-306.
- Mills, R. E. "An Evaluation of Techniques for Teaching Word Recognition," Elementary School Journal, 56 (1956), 221-225.
- Mitzel, Harold E. Teacher Effectiveness, in N. L. Gage (Ed.), Encyclopedia of Educational Research, 3rd ed., New York: MacMillan, 1960, 1481-1485.
- Morsh, J. E. "Systematic Observation of Instructor Behavior," USAF Pers. Train. Res. Cent. Development Report, 1956, No. AFPTRC-TR-56-32.
- \_\_\_\_\_, and E. W. Wilder. "Identifying the Effective Instructor: A Review of the Quantitative Studies 1900-52," Research Bulletin No. AFPTRC-TR-54-44, USAF. Personnel Training Research Center, San Antonio, Texas, 1954.
- Ober, Richard L. "The Reciprocal Category System," Journal of Research and Development in Education, 4, No. 1 (Fall, 1970), 34-51.
- Powell, E. R. "Teacher Behavior and Pupil Achievement," Classroom Interaction Newsletter, 3 (1968), 3-5.
- Ragosta, Marjorie, Robert Soar, Ruth Soar, and Linda Stebbins. "Sign versus Category: Two Instruments for Observing Levels of Thinking," Paper presented at AERA, New York, 1971.
- Robinson, Helen M. Reading Instruction: Research, in Lee Deighton (Ed.), Encyclopedia of Education, Vol. 7, New York: MacMillan, 1971, 406-412.

Rosenshine, Barak. "Interpretive Study of Teaching Behaviors Related to Student Achievement," Final Report. September 1970a, ED 051 116.

\_\_\_\_\_. "Evaluation of Classroom Instruction," Review of Educational Research, 40 (1970), 2, 279-301.

\_\_\_\_\_. Teaching Behaviors Related to Pupil Achievement: A Review of Research, in Ian Westbury and Arno A. Bellack, (Eds.), Research into Classroom Processes, New York: Teachers College Press, 1971a, 51-98.

\_\_\_\_\_, and Norma Furst. Research on Teacher Performance Criteria, in B. O. Smith (Ed.), Research in Teacher Education: A Symposium, Englewood Cliffs, New Jersey: Prentice-Hall, 1971b, 137-172.

\_\_\_\_\_, and Norma Furst. The Use of Direct Observation to Study Teaching, in R. M. W. Travers (Ed.), Second Handbook of Research on Teaching, Chicago: Rand McNally, 1973, 122-183.

Schantz, B. M. B. "An Experimental Study Comparing the Effects of Verbal Recall by Children in Direct and Indirect Teaching Methods as a Tool of Measurement," Dissertation, Pennsylvania State University, 1963.

Science Research Associates (SRA) Achievement Test Series, Chicago: Science Research Associates, Multilevel Reading Test, Forms C and D, 1963. Primary II, Forms E and F, 1972.

Shuy, Roger W. "Sociolinguistic Strategies for Studying Urban Speech," Viewpoints, Bulletin of the School of Education, Indiana University, 47 (March, 1971), 2, 1-25.

Siegal, M. A., and Barak Rosenshine. "Teacher Behavior and Student Achievement in the DISTAR Program," Chicago Principals Reporter, 62 (1972), 24-28.

Soar, Robert S. "Methodological Problems in Predicting Teacher Effectiveness," Journal of Experimental Education, 32 (1964), 287-291.

\_\_\_\_\_. "An Integrative Approach to Classroom Learning," Final Report, Public Health Service Grant No. 5-R11 MH 01096, and National Institute of Mental Health Grant No. 7-R11-MH 02045, Philadelphia: Temple University, 1966, ED 033 749.

- Soar, Robert S. Pupil Needs and Teacher-Pupil Relationships: Experience Needed for Comprehending Reading, in Amidon and Hough (Eds.), Interaction Analysis: Theory, Research and Applications, Reading, Massachusetts: Addison-Wesley, 1967, 243-255.
- \_\_\_\_\_. "Optimum Teacher-Pupil Interaction for Pupil Growth," Educational Leadership Research Supplement, 26 (December, 1968), 3, 275-280
- \_\_\_\_\_. "Research Finding from Systematic Observation," Journal of Research and Development in Education, 4 (1970), 1, 116-122.
- \_\_\_\_\_, and Ruth Soar. An Empirical Analysis of Selected Follow-Through Programs: An Example of a Process Approach to Evaluation, in Ira J. Gordon (Ed.), Seventy-first Yearbook of the National Society for the Study of Education, Part II, 1972, 229-260.
- \_\_\_\_\_. "Follow-Through Classroom Process Measurement," Gainesville, Florida: University of Florida, 1976, ED 113 288.
- Solomon, D., et al. "Teaching Styles and Learning," Chicago: Center for the Study of Liberal Education of Adults, 1963.
- Spaulding, Robert L. "Achievement, Creativity, and Self-Concept Correlates of Teacher-Pupil Transactions in Elementary Schools," Cooperative Research Project No. 1352. Hempstead, New York: Hofstra University, 1965, ED 024 463.
- Stake, Robert E. "The Countenance of Educational Evaluation," Teachers College Record, 68 (April, 1967), 523-540.
- Thompson, G. R., and N. C. Bowers. "Fourth Grade Achievement as Related to Creativity, Intelligence, and Teaching Style," AERA Paper, Chicago, 1968.
- Torrance, E. Paul, and Elaine Parent. "Student Attitude Inventory in School Mathematics Study Group (SMSG)," in Characteristics of Mathematics Teachers that Affect Students' Learning, 1966, ED 010 378.
- Travers, Robert M. W. Some Further Reflections on the Nature of a Theory of Instruction, in Ian Westbury and Arno A. Bellack (Eds.), Research into Classroom Processes: Recent Developments and Next Steps. New York: Teachers College Press, 1971, 23-47.

- United States Office of Education (USOE). "Planning/Evaluation Study: Impact of Educational Innovation on Student Performance," Executive Summary. USOE/Office of Planning, Budgeting, and Evaluation, November, 1976 (LONGSTEP).
- Walberg, Herbert J., and Gary L. Anderson. "Classroom Climate and Individual Learning," Journal of Educational Psychology, 59 (1968), 414-419.
- Wallen, N. E. "Relations Between Teacher Characteristics and Student Behavior: Part III," Cooperative Research Project No. 2628, USOE, Salt Lake City: University of Utah, 1966.
- Wang, Margaret C. (Ed.). "The Use of Direct Observation to Study Instructional-Learning Behaviors in School Settings," AERA Symposium, New Orleans, 1973, ED 100 798.
- Webb, J. N. "Taxonomy of Cognitive Behavior," Journal of Research and Development in Education, 4 (1970), 23-33.
- Withall, J. "Development of a Technique for the Measurement of Socio-emotional Climate in Classrooms," Journal of Experimental Education, 17 (1949), 347-361.
- Wood, Samuel. "A Multidimensional Model for the Observation, Analysis, and Assessment of Classroom Behavior," Journal of Research and Development in Education, Vol. 4 (Fall, 1970), 1, 84-102.

## APPENDIX A

## APPENDIX A

### DESIGN OF THE ELEMENTARY READING IMPROVEMENT PROGRAM

The Reading Improvement Program for East Baton Rouge Public Schools, now in its third year of implementation, has been selected by the American Institutes for Research for recommendation to the National Right to Read Program of the United States Office of Education for national dissemination. Notice of this honor was received by Edna West, designer of the program, at East Baton Rouge Parish School Board on October 10, 1974.

The Reading Improvement Program, which is one of several reading programs operative in East Baton Rouge, is supported by local funding, and receives no support from agencies outside the parish. This program originated as a result of a questionnaire which was sent to every teacher in the parish. The results of that questionnaire were tabulated and the program built upon those requests which were most often made by elementary teachers. These requests included relevant in-service training for teachers, multi-level materials, help in meeting individual needs, opportunities to observe and communicate with other teachers, and opportunities to work with reading experts.

The extensive in-service program for teachers is designed to improve reading instruction through teacher training

and to provide adequate learning materials. Much time is devoted to activities which will help the teachers diagnose reading difficulties and design reading programs around the needs of each child. Major efforts in teacher training and materials are concentrated in center schools, each of which serves as a resource for a group of cluster schools. In-service meetings focus on approaches to reading instruction, effective teaching techniques, and proper, effective utilization of materials and equipment. The supervisor and helping teacher work in each school on a regular basis, sometimes with individual teachers, sometimes with a class of children and a small group of teachers, sometimes in small conferences, and sometimes with an entire faculty. The principal is a part of the planning group and facilitates the program; the librarian is considered indispensable as a resource person. Parents participate by relieving teachers and by participating in classroom activities. Teachers from cluster schools come to center schools or other cluster schools to work with the supervisor, helping teacher and/or classroom teachers. The purpose of these visitations is to provide the teacher with an understanding of the needs of the learner, to make available as much interesting and varied material as possible, and to help the teacher develop skills to create, in his/her own classroom, successful approaches to the teaching of reading. Follow-up work in the cluster schools is accomplished by classroom teachers and helping reading teachers. Workshops

planned and executed by classroom teachers with the help and guidance of the helping reading teachers and supervisor, are an important element of the program, since they give each participant an opportunity to be an integral part of the process.

--Report of program design by Edna West



## APPENDIX B

Observation Checklist for a Multi-approach Individualized Elementary Reading Program in a Self-contained Classroom as Modified in Validation Study at Louisiana State University, October 16, 1975

School \_\_\_\_\_ Classroom \_\_\_\_\_ Date \_\_\_\_\_ Observer \_\_\_\_\_ Total \_\_\_\_\_

I. Variety of approaches to reading instruction

1 2 3 Tot

A. Language experience approach

1. Pupils are provided real and/or vicarious experiences to stimulate verbal responses
2. Pupils have class word lists in view
3. Pupils have individual word lists or collections
4. Pupils use word collections in 2 and 3 above  
e.g., matching words with similar phonetic or structural parts or meanings, arranging alphabetically or by categories, making into glossaries, forming sentences or stories
5. Pupils write or dictate their ideas
6. Pupils read their own and other pupils' writings

B. Individualized reading approach

7. Pupils have available a wide variety of reading material  
e.g., Scholastic, library, and trade books, basal texts, etc.
8. Children read books of their own choosing
9. Children are provided with independent reading time to read at their own pace
10. Children do creative, independent level follow-up activities for the books they read
11. Children have individual reading conferences with teacher

C. Basal text approach-modified for individual needs

12. Basal texts are used as individualized reading material
13. Basal manuals and/or texts are used for skill group teaching
14. Basal texts are used as reference books for content area information
15. Basal texts of different levels are used
16. Various basal texts at same level are used

III. Diagnosis and development of skills

A. Oral language

17. Activities or games are used to model, stimulate, and reinforce use of standard English sentence patterns to express pupils' own ideas  
e.g., story-telling, question-and-answer games, or use of Instant Readers, Sounds of Language Series, etc.
18. Performer-audience situations are provided  
e.g., role play, Echo Plays, Story Plays, etc.

## Modified observation checklist, p. 2

	1	2	3	Tot
19. Pupil-teacher and pupil-pupil communication is encouraged				
20. Time is provided for meaningful oral sharing, planning, discussion, debates, etc.				
B. Word attack and comprehension diagnosis and achievement				
21. Informal reading inventory is administered and utilized in planning e.g., Silvaroli, etc.				
22. Diagnostic screening and/or checklists are used to identify skills which need to be taught e.g., Scholastic ditto masters, Fountain Valley screening, etc.				
23. Individual skill achievement profile is kept current for each pupil as a record of his achievement in auditory and visual discrimination skills, phonetic and structural analysis, vocabulary development, and comprehension and study skills				
C. Various means used to develop skills				
24. Specific skill is taught to the small group of children ready for it at that time				
25. Reinforcing games and activities are provided for children to use at their level of independent ability				
26. Basal reader lessons are used to teach needed skills in contextual setting				
III. Classroom interaction				
A. Pupil behavior				
27. Children ask questions openly seeming confident that they will receive supportive help				
28. Children seem absorbed in their work and appear to be self-directed (not waiting for teacher direction)				
29. Children readily help each other, cooperate easily in a peer group, and avoid disturbing other children at work				
30. Children appear to enjoy reading and learning e.g., want to continue working or reading at recess or other break, eager to explain progress or plans, to show creative product, and/or to share discoveries of "how," "how come," or "why"				
B. Teacher behavior				
31. Teacher fosters active pupil participation e.g., pupil is involved in manipulating, questioning, discussing, and figuring out how concrete objects and materials work (not just memorizing or learning from a book or lecture)				
32. Teacher allows pupil to express self freely e.g., asks, accepts, and extends pupil suggestions; pupil opinion, point of view, or question is considered with respect				

## Modified observation checklist, p.3

	1	2	Total
33. Teacher individualizes pupil's work e.g., leads pupil to question which challenges him, organizes learning around pupil's own problem or interest, has pupil work independently on what concerns him, has different pupils work at different tasks			
34. Teacher encourages openness and extension of the range of ideas generated by pupils e.g., involves pupil in open-ended situation, asks questions that require inference beyond study of the lesson			
35. Teacher encourages critical thinking of pupil to analyze and evaluate his own ideas and work e.g., asks pupil to support answer or opinion with evidence, asks pupil to predict about the unknown on the basis of the known; questions misconceptions, faulty logic, unwarranted conclusions, and comparative value of answers to develop pupil judgment, encourages pupil to put his ideas to a test, allows pupil time to think through his plan or ideas, asks pupil to evaluate his own work			
36. Teacher provides materials and tools for pupil to select and pursue subject matter e.g., makes a wide range of subject matter available, guides pupil in research skills as to collect and analyze pupil's own subject matter or to discover and correct factual errors and inaccuracies			
37. Teacher motivates pupil with intrinsic value of ideas or activities rather than tangible rewards or grades (or following concrete reinforcement needed by some reluctant readers at first, gradually moves to intrinsic motivation)			
Grand Total			

## SCORING THE OBSERVATION CHECKLIST

An observation consists of three equal time intervals. During each time interval, a tally is made in the appropriate column for each behavior observed. In a given time interval, a behavior gets only one tally mark, no matter how often the behavior occurs. Before the time interval ends, the observer should ascertain that each type of behavior observed has been checked. After the third time interval is completed, the total score for each item is entered in the Total column. (An item can have a total score of 0, 1, 2, or 3.) The total for all behaviors during the entire observation is then recorded.

Classroom Tally Sheet for ERI<sup>2</sup> Observation Checklist

		School _____	Date _____
		Classroom _____	Grand Total _____

I. Variety of approach	1 2 3 Tot	
A. Language experience		
1. exprnc to stim vrbl resp		
2. have class word lists		
3. p have own word lists		
4. p use word lists		
5. p write, dictate ideas		
6. p read own-other p ideas		
E. Individualized reading		
7. wide vrtty rdg mtrl		
8. ch read bks of own chng		
9. ch read bks at own pace		
10. do indpdnt flwup actvty		
11. indvdl p-t rdg confrnc		
C. Basal approach--modified		
12. as indvdlzd rdg mtrl		
13. manl-text for skl tchg		
14. for content informtn		
15. diffrent levels used		
16. diffrent series used		
II. Diagnosis-development/skills		
A. Oral language		
17. practc standrd English		
18. prfrmr-audnc situation		
19. p-t & p-p comctn		
20. oral sharng, plng		

B. Word attack-comprehension	1 2 3 Tot	
21. use IRI		
22. do diagnstc scrng		
23. indvdl skill profile		
C. Means to develop skills		
24. sm1 grp skl tchg		
25. reinfrcg actvts		
26. basal--skls in cntxt		
III. Classroom interaction		
A. Pupil behavior		
27. ch confdnt of suprt		
28. ch absrbd, slf-drcd		
29. ch help ch, cooperate		
30. ch enjoy rdg & lng		
B. Teacher behavior		
31. t fosters p partcptn		
32. t alws p to xprs slf		
33. t indvdlzs p work		
34. t encrgs rng of p ideas		
35. t encrgs p crtcl thnkg		
36. t prvds/p prsu/sbjt mtr		
37. t mtvts/valu of ideas		
Observer _____	Grand Total	

## APPENDIX C

APPENDIX C

COMMUNICATIONS WITH ELEMENTARY READING  
IMPROVEMENT PROGRAM PERSONNEL

215 Stanford Avenue  
Baton Rouge, La. 70808  
June 28, 1975

Dear Mrs. West,

As a research project, and as an additional evaluation of the Elementary Reading Improvement Program of East Baton Rouge Parish, correlation of pupil reading gains with degree of teacher implementation of the program would be of interest. As a means of rating the degree of teacher implementation, an observation checklist could be used by the helping reading teachers in the program to record their classroom observations as they work with their teachers. Your approval is requested to pursue this project.

As the designed and director of this program, you are best qualified to judge if the checklist seems to indicate the theoretical principles and educational objectives which should be evidenced, and whether it would produce an accurate measure of the use of the types of materials and techniques you consider to be critical to this program. I would appreciate your evaluation of the acceptability of each item in the proposed checklist. Does this checklist accurately reflect the design and spirit of the Elementary Reading Improvement Program? Your general reactions and your specific comments on any items that are not in acceptable form, and any that you think should be deleted or added would be of great help.

In your reply, please sign your name and professional position as you wish it to appear in the credits of the research report.

Thank you for your valued opinion and for your kindness.

Sincerely yours,



Patience W. Keisler



PH. (504) 828-2790

*East Baton Rouge Parish School Board*

ROBERT J. AERTKER, SUPERINTENDENT

P. O. BOX 2810

*Baton Rouge, Louisiana 70801*

July 3, 1975

L. L. MINES, PRES  
2500 TIMBER COVE DR.  
BATON ROUGE, LA. 70808

DONALD D. HUNT, VICE-PRES  
P. O. BOX 193  
ZACHARY, LA. 70791

MRS. WALLACE F. ARMSTRONG  
7922 BOCADE BLVD.  
BATON ROUGE, LA. 70808

J. O. CLAUDEL  
RT. 2 - BOX 212  
BAKER, LA. 70714

A. BRIDGER EGLIN  
P. O. BOX 1791  
BATON ROUGE, LA. 70821

RANDALL GOODWIN  
1124 WEST LAKEVIEW  
BATON ROUGE, LA. 70810

MARTIAL J. LAFLEUR, JR.  
1819 CAMELIA AVENUE  
BAKER, LA. 70714

LAWRENCE E. MOCH, SR.  
1428 N. 28TH STREET  
BATON ROUGE, LA. 70802

T. H. MONTGOMERY  
ROUTE 4  
BATON ROUGE, LA. 70808

BEN H. PEABODY  
2274 N. APTON PARKWAY  
BATON ROUGE, LA. 70805

W. HARRY PERKINS  
P. O. BOX 24974  
BATON ROUGE, LA. 70808

W. W. WELLS  
1322 BORREL AVENUE  
BATON ROUGE, LA. 70801

Mrs. Patience W. Keisler  
215 Stanford Avenue  
Baton Rouge, Louisiana 70808

Dear Patience:

I think your research project should prove interesting and valuable. You have my consent and approval to pursue it in the schools where the Elementary Reading Improvement Program is operative. I believe you must have Dr. Smiley's approval also.

I have suggested a few changes which you may or may not want to follow. As to your request that we discuss the program and your project, you know that I am available. Please call and let Helen know when you can come. Good luck!

Sincerely,

Edna West  
Elementary Supervisor

EW/hf

Attachment



215 Stanford Avenue  
Baton Rouge, La. 70808  
June 28, 1975

Dear Mary Ellen, (Molly,)

As a research project, and as an additional evaluation of the Elementary Reading Improvement Program of East Baton Rouge Parish, correlation of pupil reading gains with degree of teacher implementation of the program would be of interest. As a means of rating the degree of teacher implementation, an observation checklist could be used by the helping reading teachers in the program to record their observations as they work with their teachers. Your help with this project is requested.

As a helping reading teacher who has been involved in developing and encouraging implementation of this program for the past two years, you are well qualified to judge if the checklist seems to indicate the theoretical principles and educational objectives which should be evidenced, and whether it would produce an accurate measure of the use of the types of materials you consider to be critical to this program. I would appreciate your evaluation of the acceptability of each item in the proposed checklist. Does this checklist accurately reflect the design and spirit of the ERIP? Your general reactions and your specific comments on any items that are not in acceptable form, and any that you think should be deleted or added would be of great help.

In your reply, please sign your name and professional position as you wish it to appear in the credits of the research report.

Thank you for your valued opinion and for your kindness.

Sincerely yours,

A handwritten signature in cursive script that reads "Patience W. Keisler".

Patience W. Keisler



*East Baton Rouge Parish School Board*

ROBERT J. AERTKER, SUPERINTENDENT

P. O. BOX 1880

*Baton Rouge, Louisiana 70801*

PM. (504) 926-2780

October 1, 1975

Mrs. Patience Keisler  
215 Stanford Avenue  
Baton Rouge, Louisiana

Dear Patience,

I am hopeful that the changes in your Observation Checklist which resulted from our several discussions will bring it closer to being the useful instrument that we have been envisioning.

I am most anxious to see it reach its final form for I feel it will be of great benefit to our program. In the first place, it will give us a concrete tool to use with teachers who are wanting to make changes, but may be unsure of the direction they wish to take. Not only will it help us to make more specific suggestions when asked, but it will allow the teacher the opportunity for a continuous self-evaluation.

Secondly, your research correlating the degree of implementation of individualization with achievement test scores of children in our program will give us an additional way to evaluate the Elementary Reading Improvement Program and its effect on teachers and students, as well as provide some insights into how well the SRA Achievement Tests are measuring what we are trying to do.

If I can be of any further direct assistance, please do not hesitate to call on me.

Sincerely,

*Molly Newkome*

Molly Newkome

Instructional Reading  
Consultant

MN:jab

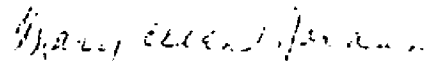
1485 Cloverdale Avenue  
Baton Rouge, Louisiana  
July 28, 1975

Dear Patience,

This letter is in reply to the letter you sent me re: the checklist you designed as a means of rating the degree of teacher implementation of the Elementary Reading Improvement Program. After very careful study and several conferences with you about it, I sincerely feel that the checklist does accurately reflect the design and spirit of E.R.I.P. I hope the comments I have made to you will be beneficial in the use of the checklist.

I know that this checklist will have a two-fold purpose even though you did not design it with this in mind. I feel that it will make me do a better job in working with my teachers. If I can be of any further help to you please do not hesitate to call me.

Sincerely,



Mary Ellen Jordan  
Instructional Consultant  
East Baton Rouge Parish Schools

mej

## APPENDIX D



		ANSWERS TO QUESTIONS			
		A		B	
		YES	NO	YES	NO
21.	Pupil-teacher and pupil-pupil communication is encouraged.				
22.	Time is provided for meaningful oral sharing, planning, discussion, and debates, etc.				
B. Word attack and comprehension					
23.	Informal reading inventory is used (as Silvercili, etc.)				
24.	Diagnostic screening is used (as Scholastic ditto masters, Fountain Valley screening, etc.)				
25.	Individual skill achievement profile is kept current for each pupil				
	Various means are used to develop skills:				
26.	a. Small group teaching				
27.	b. Reinforcing games and activities				
28.	c. Basal reader lessons with skills in context				
29.	Checklists and specific materials are used for diagnosing and teaching visual and auditory discrimination skills when needed(as Target Red)				
III. Classroom interaction					
A. Pupil behavior					
30.	Children ask questions openly seeming confident that they will receive supportive help				
31.	Children seem absorbed in their work and appear to be self-directed (not waiting for teacher direction)				
32.	Children readily help each other, cooperate easily in a peer group, and avoid disturbing other children at work				
33.	Children appear to enjoy reading and learning				
B. Teacher behavior					
34.	Teacher fosters active pupil participation				
35.	Teacher allows pupil to express self freely				
36.	Teacher individualizes pupil's work e.g., leads pupil to question which challenges him, organizes learning around pupil's own problem or interest, has pupil work independently on what concerns him, has different pupils work at different tasks.				
37.	Teacher encourages openness and extension of the range of ideas generated by pupils. e.g., involves pupil in open-ended situation, asks questions that require inference beyond study of lesson.				
38.	Teacher encourages critical thinking of pupil to analyze and evaluate his own ideas and work e.g., asks pupil to support answer or opinion with evidence, asks pupil to predict about the unknown on the basis of the known, questions misconceptions, faulty logic, unwarranted conclusions, and comparative value of answers to encourage pupil judgment, encourages pupil to put his ideas to a test, allows pupil time to think through his plan or ideas, asks pupil to evaluate his own work				
39.	Teacher provides materials and tools for pupil to select and pursue subject matter e.g., makes a wide range of subject matter available, guides pupil in research skills as to collect and analyze pupil's own subject matter or to discover and correct factual errors and inaccuracies				

ANSWERS TO QUESTIONS			
A		B	
YES	NO	YES	NO

40. Teacher evaluates pupil work on basis of each individual's own progress and ability level
41. Teacher motivates pupil with intrinsic value of ideas or activities rather than tangible rewards or grades

Over-all comments:

- Which of the above items do you think should be deleted from this checklist in order to improve the reading-language arts program represented? (List item numbers; comments welcome)
- What elements important to an individualized elementary reading-language arts program do you think have been left out of this checklist?
- Other reactions to this observational checklist:

## APPENDIX E



## APPENDIX E

### SCHOOL BOARD PERMISSION

215 Stanford Avenue  
Baton Rouge, La. 70808  
September 30, 1975

Mr. Robert J. Aertker, Superintendent  
East Baton Rouge Parish School Board Office  
Post Office Box 2950  
Baton Rouge, Louisiana 70821

Dear Mr. Aertker,

This letter requests permission to conduct the research for a dissertation in the schools using the Elementary Reading Improvement Program of the East Baton Rouge School Board. The study, which will be an evaluation of this reading program, may prove of value to the school board staff also. The study will attempt to assess the relationship between classroom process (teaching-learning behaviors) and educational product (pupil achievement). Surprisingly little of this vital sort of research has been done due to problems related to observing in classrooms. I think that data for this evaluation could be collected by the three instructional consultants in the program as they carry out their regular classroom visits. The form of recording observations would be more structured than it has been heretofore.

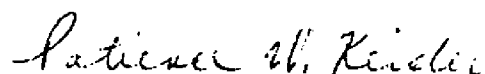
The title of the study I propose is "A Study of the Relationship of Pupil Achievement to the Degree of Teacher Implementation of an Individualized Reading Program." A random sampling of teachers will be chosen from among the center and cluster school teachers involved in the program. Their classrooms will be rated on the basis of the enclosed observation checklist to determine the degree to which specific elements of the program seem to be employed. Each of the three instructional reading consultants in the program would observe and rate the selected teachers with whom she normally works. Use of the checklist would constitute a more specific record of the usual observations made in preparation for conferences with the classroom teachers. High scores on the checklist should indicate classrooms with high implementation of the reading program. Pupil gains in reading achievement of the classrooms showing highest implementation of the

program will be compared with classes showing lowest implementation as a means of determining effectiveness of the Elementary Reading Improvement Program. Achievement gains will be measured by using a class mean difference between pre- and posttest scores. I would need your permission to use the SRA test scores.

Information regarding individual teachers or pupils will be kept confidential and used only as statistical data. No publication of findings will be made (other than the dissertation) without permission from the East Baton Rouge Parish School Board Office. I will provide you a copy of the study when it is completed.

Thank you for your consideration.

Yours truly,

A handwritten signature in cursive script, reading "Patience W. Keisler".

Patience W. Keisler

checklist encl.



*East Baton Rouge Parish School Board*

140

ROBERT J. AERTKER, SUPERINTENDENT

P. O. BOX 2950

*Baton Rouge, Louisiana 70821*

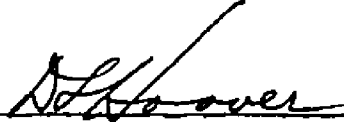
December 15, 1975

Memo to: Mrs. Patience Keisler  
From: Donald L. Hoover, General Coordinator  
Subject: Reading Study Project

I have reviewed your request to conduct a study in the area of reading in some elementary schools of our parish in the Reading Program. The Superintendent has informed me that we will cooperate with you in the study. Please let this letter serve as your authorization.

I would suggest that a copy of this letter should be on hand should the study be questioned by the principal of the school involved.

If you have any further question or need any assistance in this matter, please feel free to contact me.

  
\_\_\_\_\_  
Donald L. Hoover  
General Coordinator

DLH/mmg  
cc: Mr. Aertker  
Mr. Thom

## APPENDIX F



*East Baton Rouge Parish School Board*

ROBERT J. AERTKER, SUPERINTENDENT

P. O. BOX 2880

*Baton Rouge, Louisiana 70821*

PH. (504) 928-2790

September 2, 1975

Memo to: Principals and Teachers of Center and Cluster Schools  
East Baton Rouge Parish Elementary Reading Improvement Program

From: Mary Ellen Jordan, Patience Keisler, and Molly Newkome  
Instructional Consultants

Subject: Testing (WRAT and SRA)

Testing in the Center and Cluster Schools will begin during the week of September 8, 1975. The following points are for your information.

Grades 1 and 2:

The Instructional Consultants will administer the WRAT in grades 1 and 2 in the Center and Cluster Schools to randomly selected students. This testing will begin Monday, September 8, 1975 and should be completed in two weeks. You will be contacted concerning the dates of the testing in your school.

Grades 3 and 4:

As in the past, teachers are asked to administer and score the SRA Primary II Reading Test, Form E. Put the results of the test on the enclosed Class Record Forms. List your boys in alphabetical order on one sheet and the girls in alphabetical order on another. Put the last name first. Return Class Record Forms and test booklets to your school office by September 19, 1975.

Grades 5 and 6:

Teachers in grades 5 and 6 will administer the SRA Multilevel Reading Test, Form C, using the IBM answer sheets provided by Data Processing. You may have some answer sheets for children no longer in your class. Please destroy them as they cannot be used for any other child. For any child who does not have a printed form, use a blank answer sheet and print the child's name and student number on it. Return these answer sheets to Data Processing at the Central Office by Friday, September 12, 1975 where they will be machine scored. The answers for the reading test should be recorded only in blanks 1 - 92 on the blue answer sheet.

Your cooperation in this important testing program is essential and is greatly appreciated.

MEJ/PK/MN:hf

Approved:

  
Mrs. Edna West, Elementary Supervisor

cc: Dr. Hoover  
Mr. Thom



*East Baton Rouge Parish School Board*

ROBERT J. AERTKER, SUPERINTENDENT

P. O. BOX 2930

*Baton Rouge, Louisiana 70821*

April 20, 1976

PH. (504) 926-2780

Memo to: Principals and Teachers in East Baton Rouge Parish  
Elementary Reading Improvement Program

From: Mary Ellen Jordan, Patience Keisler, Molly Newkome  
Instructional Consultants

Subject: Post-Testing (WRAT and SRA)

Post-testing in the Elementary Reading Improvement  
Program will be according to the following schedule.

Grades 1 and 2:

The WRAT (Wide Range Achievement Test) will be administered in grades 1 and 2 to the same children who were pre-tested in the fall. This testing will begin Monday, May 3, 1976 and should be completed in two weeks. You will be contacted concerning the dates of the testing in your school.

Grades 3 and 4:

Teachers in grades 3 and 4 will administer and score the SRA Reading Test, Form F. A scoring key and a conversion table have been included in your packet of tests. Please return these with your Class Record Form. The test results will be recorded on the same Class Record Form used for Form E reading test in the fall. In the same envelope with your tests you will receive a copy of the Class Record Form which you turned in along with your tests in the fall. Space has been provided on this form for you to enter the Form F reading test results. Do not add any students who were not on the original list. Draw a line through names of students on this Class Record Form who do not take the SRA Form F Reading test. (Your Class Record Form should include only pupils who took both tests, Form E in September, 1975 and Form F in May, 1976.)

In the envelope which has been provided please return the tests, scoring key, conversion table and Class Record Form to your school office. This should be returned by May 14, 1976.

Grades 5 and 6:

Teachers in grades 5 and 6 will administer the SRA Multilevel Reading Test, Form D using the IBM answer sheets provided by Data Processing during the week of May 10-14, 1976. Only children for whom answer sheets are provided will be tested. You may have some answer sheets for children no longer in your class. Please destroy them as they can not be used for any other

## Page 2 Post-Testing (WRAT and SRA)

child. These tests will be machine scored by Data Processing at the Central Office. The answers for the reading test should be recorded only in blanks 1-92 on the IEM answer sheet.

Return the test booklets and IBM answer sheets in the envelopes which were provided with your name, grade, and school on each, to your school office as soon as testing is completed. They will be picked up by May 14, 1976.

Grades 3, 4, 5 and 6:

To ensure validity of test scores the SRA testing schedule in grades 3, 4, 5 and 6 should be uniform. Please use the following testing schedule:

Grades 3 and 4:

Tuesday, May 4: Reading Comprehension  
(approximate time required-  
65 minutes)

Wednesday, May 5: Reading Vocabulary  
(approximate time required-  
55 minutes)

Grades 5 and 6:

Week of May 10th: Reading Comprehension and  
Vocabulary  
(approximate time required-  
77 minutes)

Your cooperation in this most important testing program is essential and is greatly appreciated.

MEJ/PK/MN:jn

APPROVED:

  
Edna West, Elementary Supervisor

cc: Dr. Hoover

## APPENDIX G



## APPENDIX G

STATISTICAL ANALYSIS SYSTEM 17002 TUESDAY, JANUARY 18, 1977 11

TEACHER	SCHOOL	NUM	BLACK	PERBLACK	CL1	CL2	CL3	CLTOT	PRERAW	PORAW	RGAIN	
GRADE	RACE											
6	1	2	28	10	35.714	40	13	33	86	39.2500	45.4286	6.1786
5	2	4	29	0	0.000	46	14	32	92	30.8621	38.2759	7.4138
4	3	4	28	0	0.000	43	7	27	77	35.4643	41.2857	5.8214
3	4	4	25	0	0.000	41	13	30	84	27.4000	37.2800	9.8800
4	5	2	23	1	4.348	41	11	30	82	33.0435	37.3913	4.3478
6	6	2	30	6	20.000	44	12	28	84	44.6207	49.2667	5.5172
6	7	4	27	3	0.000	32	10	33	75	39.1111	45.9259	6.8148
5	9	1	28	13	46.429	7	2	0	9	27.6786	35.8214	8.1429
4	10	3	16	2	12.500	9	2	0	11	38.0625	44.5625	6.5000
6	11	5	28	0	0.000	8	4	0	12	66.3929	69.1071	2.7143
4	12	2	24	2	8.333	16	1	0	17	30.7917	36.2917	5.5000
6	13	1	28	11	39.286	13	1	1	15	40.0357	44.1071	4.0714
4	14	5	25	0	0.000	9	3	0	12	44.6400	47.2000	2.5600
6	15	5	27	0	0.000	8	2	2	12	60.4444	65.8519	5.4074
3	16	2	22	1	4.545	14	3	6	17	24.5455	37.7273	13.1818
5	17	8	21	11	52.381	41	19	32	92	23.0476	29.4762	6.4286
3	18	8	21	9	42.857	43	16	33	92	21.1905	31.7619	10.5714
6	19	10	23	5	21.739	40	16	33	89	36.3750	41.3333	4.9583
4	21	8	24	13	54.167	40	13	33	86	30.8333	37.7917	6.9583
4	22	6	20	13	65.000	39	16	33	88	30.5000	36.9500	6.4500
5	23	9	19	14	73.684	42	13	33	88	21.2632	19.7895	-1.4737
6	24	8	28	15	53.571	41	15	31	87	36.5714	41.0357	4.4643
6	25	6	22	15	68.182	6	3	2	11	33.4091	36.6818	3.2727
5	26	6	26	17	65.385	5	2	8	15	20.5000	27.7308	7.2308
6	27	9	22	19	86.364	9	5	18	32	25.7727	23.0000	2.2273
6	28	6	22	12	54.545	9	3	7	19	28.2273	37.5000	9.2727
4	29	10	19	3	15.789	9	3	12	24	37.8947	46.0000	8.1053
3	31	10	14	3	21.429	15	5	12	32	24.3571	30.3571	6.0000
4	33	15	17	10	58.824	41	17	28	86	29.1176	35.1765	6.0588
5	34	12	27	27	100.000	38	18	27	83	22.3704	23.3333	0.9630
6	35	12	17	17	100.000	33	11	26	70	16.7647	28.2941	11.5294
6	36	15	23	11	47.826	29	11	29	69	45.3913	49.4783	4.0870
6	37	14	27	13	48.148	31	4	17	52	36.3333	33.3704	-2.9630
5	38	11	24	24	100.000	30	10	24	64	15.1667	24.9583	8.7917
6	39	13	27	3	11.111	33	10	30	73	56.7778	61.6667	4.8889
3	40	13	15	0	0.000	32	9	29	70	38.6000	41.7333	3.1333
3	41	11	23	23	100.000	11	5	5	21	19.1304	32.7826	13.6522
4	42	11	24	24	100.000	5	1	1	7	22.9167	27.4167	4.5000
5	44	14	22	10	45.455	14	4	12	30	23.3182	33.0000	9.6818
4	45	12	22	22	100.000	8	5	6	19	17.7273	21.2727	3.5455
5	46	12	22	22	100.000	10	5	8	23	8.9130	24.0435	15.1304
6	47	15	22	10	45.455	16	7	10	33	44.8182	45.9091	2.0909
4	48	11	23	23	100.000	11	3	3	17	21.3478	28.0870	6.7391

146

## APPENDIX H

## APPENDIX H

### ELEMENTARY READING IMPROVEMENT PROGRAM SCHOOLS OF EAST BATON ROUGE PARISH BATON ROUGE, LOUISIANA

#### Howell Park Cluster

Dalton

Fairfields

Howell Park

Northdale

Winbourne

#### Westdale Cluster

Belfair

Dufrocq

Highland

Mayfair

Westdale

#### Park Ridge Cluster

Brownfields

Baker Heights

Bakerfield

Sherwood Forest

Park Ridge

## APPENDIX I

## APPENDIX I

Table 14

Frequency Distribution of Classrooms According to  
Mean Pretest and Mean Gain for High Implementers

Raw Score Mean Gain	0-9	10-19	20-29	30-39	40-49	50-59	60-69
15-17							
12-14		1					
9-11		1	2				
6-8			2	5	1		
3-5				5	1	1	
0-2			1				
-3--1			1	1			

Table 15

Frequency Distribution of Classrooms According to  
Mean Pretest and Mean Gain for Low Implementers

Raw Score Mean Gain	0-9	10-19	20-29	30-39	40-49	50-59	60-69
15-17		1					
12-14		1	1				
9-11			2				
6-8			4	3			
3-5		1	1	1	2		2
0-2			1		1		
-3--1							

## VITA

Patience Weidt Keisler, daughter of Bew Patten Weidt and William C. J. Weidt of Mount Vernon, New York, was born September 30, 1927, at Mount Kisco, New York. She was educated in public schools of Mount Vernon, New York, graduating from A. B. Davis High School in 1946. After attendance at Carthage College, Carthage, Illinois, and Midland College, Fremont, Nebraska, she received a Bachelor of Science degree from the latter in 1952. Majoring in elementary education and minoring in library science, she completed the Master of Education degree at Louisiana State University (L.S.U.), Baton Rouge, Louisiana, in 1972. The work for her doctoral degree included a major in elementary education with emphasis on the study of reading and a minor in anthropology emphasizing cultural and linguistic studies.

Employed by the East Baton Rouge Parish School Board of Baton Rouge, Louisiana, since 1966, she has served in the following capacities: elementary classroom teacher (seven years), instructional consultant in the Elementary Reading Improvement Program, instructional consultant in the Department of Research and Curriculum Development developing curriculum guides and procedures for the adoption of textbooks and other materials of instruction, Title I helping reading teacher, Title I program evaluator in the Department

of Research and Program Evaluation, writer on Criterion-referenced Test Writing team, presenter of ESS science units for L.S.U. course offered jointly by the parish school system.

In other professional activities, she served as graduate assistant to L.S.U. Supervisor of Student Teachers, as graduate assistant in university language arts methods course, as member of Southern Association school visitation committee, as vice-president and president of Capital Area Reading Council, as member of executive committee of the state-wide Louisiana Reading Association, as delegate to International Reading Association convention, as member of parish textbook adoption committees for elementary reading and elementary mathematics, and as member of state textbook adoption committee for elementary mathematics.

Her husband is James Edwin Keisler. They have three sons: James Edwin, Jr., William Bryan, and Paul Timothy.

## EXAMINATION AND THESIS REPORT

Candidate: Patience W. Keisler

Major Field: Education

Title of Thesis: A STUDY OF THE RELATIONSHIP OF PUPIL ACHIEVEMENT TO THE  
DEGREE OF TEACHER IMPLEMENTATION OF AN INDIVIDUALIZED  
ELEMENTARY READING PROGRAM

Approved:

Wesley M. Cookston  
Major Professor and Chairman

James B. Traynham  
Dean of the Graduate School

### EXAMINING COMMITTEE:

Dany Adams  
Lynne Long  
John E. Hurd  
Eric R. Hurston

Date of Examination:

July 30, 1979